RO-S-68-1

AD 674190

A VECTOR SPACE DERIVATION-USING DYADS-OF WEIGHTED LEAST SQUARES FOR CORRELATED NOISE

(Special Report)

by

James S. Pappas

JUNE 1968

D D C
RECEIVED
SEP 1 1 1968
RECEIVED
B

U. S. ARMY TEST AND EVALUATION COMMAND
ANALYSIS AND COMPUTATION DIRECTORATE
DEPUTY FOR NATIONAL RANGE OPERATIONS
WHITE SANDS MISSILE RANGE, NEW MEXICO

DISTRIBUTION OF THIS DOCUMENT IS UNLIMINTED

Reproduced by the
CLEARINGHOUSE
for Federal Scientific & Technical
Information Springfield Va. 22151

A VECTOR SPACE DERIVATION-USING
DYADS-OF WEIGHTED LEAST SQUARES
FOR CORRELATED NOISE

(Special Report)

by

James S. Pappas

Approved by
GUENTHER HINTZE
Director

June 1968

U. S. ARMY TEST AND EVALUATION COMMAND

ANALYSIS AND COMPUTATION DIRECTORATE
DEPUTY FOR NATIONAL RANGE OPERATIONS
WHITE SANDS MISSILE RANGE, NEW MEXICO

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

CONTENTS

ABSTRACT

Matrix-analysis and recursive matrix computing subroutines offer hope of relieving the current computer data deluge. Classical weighted least squares for multi-variable parameter estimation in the presence of correlated noise are developed in a geometrical vector space setting. Rank-one matrices, or dyads, are used extensively, especially in obtaining gradients of traces of variance matrices.

TEXT NOT REPRODUCIBLE

INTRODUCTION

This report develops the classical weighted least-squares theory in a vector-space setting. Computer programs and subroutines which operate on larger packages of data in the form of data-matrices and large arrays of system variables as Euclidean vectors offer great hope of relieving the current data deluge plague.

Our current computer programming procedures are based on arithmetic operations on algebraic field elements such as addition, multiplication, division, and integration of scalars. The state space formulation requires arithmetic units which operate on matrices as elements of an algebraic ring, vector space, etc.

In the classical weighted least squares theory one analytically and computer-wise works with tedious summation after summations of scalar variables. In the modern theory one analytically and computer-wise works with vector space theory, square and rectangular data matrices of full and non-full rank and their inverses and psuedo inverses. Computer economy in data storage and computing time are sought through the applications of clever recursive matrix numerical analysis algorithms.

This report is the second of a series developing the modern state vector recursive estimation theory. The essential areas for understanding the theory are:

1. Unweighted Least Squares Parameter-Vector Estimation and the Variance-of-the-Estimate Matrix.

2. Discrete Matrix Recursive Methods Applied to (1) for Real Time (on line) Computer Processing.

3. Weighted Least Squares Parameter Estimation and Variance-of-the-Estimate Matrix for Correlated Noise.

4. Discrete Matrix Recursive Methods Applied to (3) for Real Time Computer Brocessing.

5. Recursive Weighted Least Squares State-Vector Estimation Theory (Kalman Theory).

Item (1) and (2) are completed and published in reference (4). Item (3) is the contents of the current report. Items (4) and (5) are near completion.

NOTATION

The notations used in the report is an effort to blend the notation of Friedman for inner-products and dyadic products with the current journal-literature on vector-spaces, psuedo-inverses, state-vectors, etc.

$\underset{p\text{x}k}{X}$	capital letters designate matrices of size p rows and k columns.
$x(k\rangle$:	when p = 1, the matrix is called a column vector, and we use Friedmans symbol to distinguish this matrix.
$\langle p(x$:	when k = 1, the matrix is a row vector of dmension p.
$\langle p)x\ y(p\rangle$	"inner-product" or scalar product of two vectors.
$y(p\rangle\langle p)x$:	"outer-product" or dyadic product of two vectors.

$\underset{p\text{x}k}{X} = [\,x(p\rangle_1, \ldots, x(p\rangle_k\,]$. Matrix X partitioned into a row k-tuple of column vectors from a p-space.

$$\underset{p\text{x}k}{X} = \begin{bmatrix} \overset{1}{\langle k)}x \\ \vdots \\ \overset{p}{\langle k)}x \end{bmatrix}$$

Matrix X partitioned into a p-column tuple of row vectors from a k-space.

x	small x is a scalar
x^i	scalar from a column vector
x_j	scalar from a row vector

Scalar here is a "real field" element.

SECTION 1. PRELIMINARY DISCUSSION

Consider the system of two vector equations

$$x(k+1)\rangle = \Phi(k+1, k)x(k)\rangle + f(k)\rangle + u(k)\rangle \tag{1}$$

and

$$z(k)\rangle = \underset{mxp}{H(k)}x(k)\rangle + v(k)\rangle \tag{2}$$

where:

$x(k)\rangle$, $x(k+1)(p\rangle$ are p-dimensional column vectors describing the states at stage k and stage k+1.

$\Phi(k+1, k)$ is a pxp state transition matrix.

$f(k)(p\rangle$ is a p-dimensional deterministic forcing vector for which we can write a vector function.

$u(k)(p\rangle$ is a p-dimensional uncertainty or noise vector, it is the composite of the random noises and the variables we fail to model.

$z(k)(m\rangle$ is the m-dimensional observation vector, m is less than or equal to p.

$\underset{mxp}{H}(k)$ is the known matrix describing how the state vector is functionally related to the observation vector (if the instruments were noise free).

$v(k)(m\rangle$ is an m-dimensional additive instrument noise vector.

The special case of

$$f(k)\rangle = u(k)\rangle = 0\rangle \tag{3}$$

and

$$\Phi(k+1, k) = I \tag{4}$$

and

$$H(k) = H_0 \equiv \text{a constant matrix yields} \tag{5}$$

$$\begin{aligned} x(2)\rangle &= I\, x(1)\rangle \\ x(3)\rangle &= I\, x(2)\rangle = x(1)\rangle \\ &\;\vdots \\ x(k)\rangle &= x(1)\rangle \quad \text{for all } k. \end{aligned} \tag{6}$$

And

$$z(k)(m)\rangle = \underset{m \times p}{H_0}\, x(1)(p)\rangle + v(k)(m)\rangle \tag{7}$$

Define the vector

$$\alpha(m)\rangle_0 = \underset{m \times p}{H_0}\, x(1)(p)\rangle \tag{8}$$

and

$$z(k)\rangle = \alpha\rangle_0 + v(k)\rangle. \tag{9}$$

The block diagram of equation (9) is

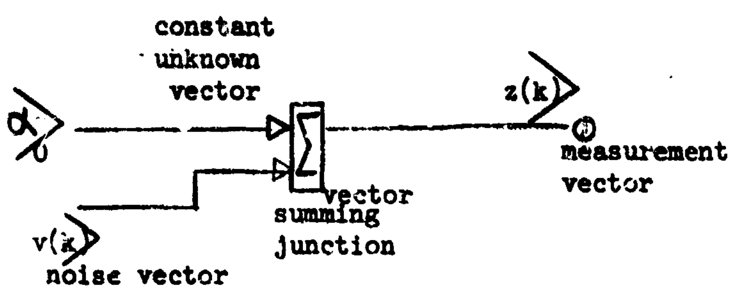

Fig. 1 Block of Vector Summing Junction

The block diagram of Equation (7) is

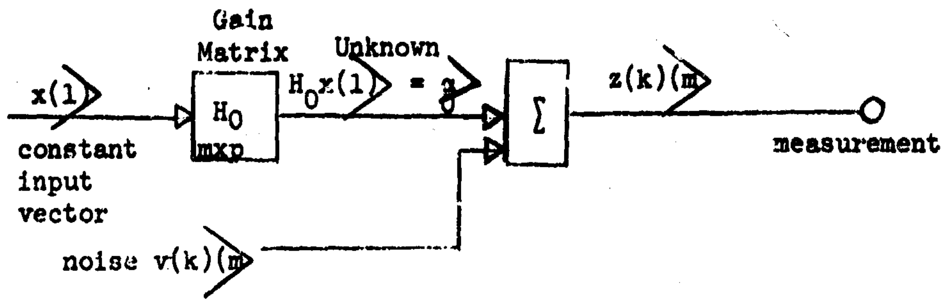

Fig. 2 Block Diagram of Eq. 7 as Device with Matrix Gain plus Additive Noise

The graph of equation (9) is a random dispersion about a constant vector in m-space as shown in Figure 3.

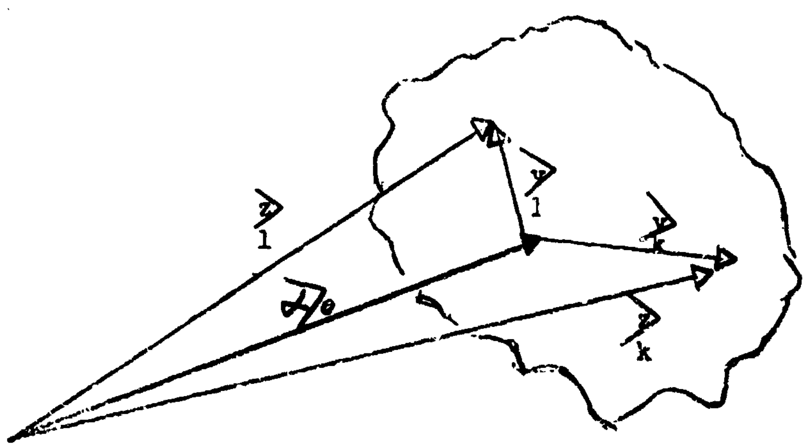

Fig. 3 Graph of Eq. 9 k-Noisy Vectors in M-Space About a Point

The graph of equation (7) is shown as a transformation on a constant vector x(1)> in p-space to a sub-space of dimension m plus an additive m-dimensional noise vector in Figure 4.

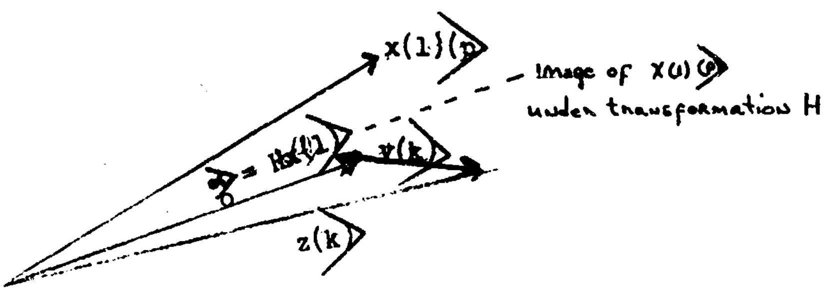

Fig. 4 Graph of Equation 7

Noise Free Conditions

The noise-free condition for the multivariable case is discussed merely to motivate some algebraic concepts to compare with the almost trivial scalar case.

When the noise is zero in Fig. 3 and Eq. 9 we have

$$z(k)\rangle = \alpha_0\rangle \qquad (10)$$

hence one measurement of $z(1)\rangle$ is adequate to find $\alpha_0\rangle$.

Since $\alpha_0\rangle$ by Eq. 8 has factors

$$\alpha_0(m\rangle = \underset{m\times p}{H_0}\, x(1)(p\rangle \qquad (11)$$

two interpretations of interest can occur.

Interpretation I

Input Measurement.

The m-dimensional vector equation (11), when $\alpha_0\rangle$ is a known m vector and $\underset{m\times p}{H_0}$ is a known mxp "gain" matrix, presents the problem to solve for the p-dimensional input vector $x(1)(p\rangle$ where $p \geq m$.

When $p = 1$, that is the scalar case

$$\alpha_0 = h_0 \, x(1) \tag{12}$$

hence

$$h_0^{-1} \, \alpha_0 = x(1). \tag{13}$$

The scalar h_0 has an inverse, however the mxp matrix H_0 does not have a conventional inverse except when $p = m$ and H is full rank; when p is greater than m the psuedo-inverse is a valuable tool to obtain <u>part of the solution</u>.

<u>Interpretation II</u>.

<u>Instrument Gain Calibration</u>.

The second case of interest for the noise free case is when $\alpha\rangle_0$ is known and we know the inputs, then the problem is to solve for the gain matrix. We have

$$\alpha(m\rangle_0 = \underset{mxp}{H} \, x(1\rangle = z\rangle_k \tag{14}$$

In equation (14) we have 1 vector equation (or m scalar equations) with mxp unknowns. If we use p different known inputs then

$$\begin{aligned} z\rangle_1 &= \underset{mxp}{H} \, x(1\rangle_1 = \alpha\rangle_{01} \\ z\rangle_2 &= H \, x(1\rangle_2 = \alpha\rangle_{02} \\ &\;\;\vdots \\ z(m\rangle_p &= H \, x(1\rangle_p = \alpha\rangle_{0p} \end{aligned} \tag{15}$$

or packaging the data as an mxp matrix

$$\left[z(m\rangle_1 \, , \; z(m\rangle_2 \, , \; \ldots \; z(m\rangle_p \right] = \underset{mxp}{Z} \tag{16}$$

and

$$Z = \left[H \, x(1\rangle_1 \, , \; H \, x(2\rangle_2 \; . \; . \; . \; , \; H \, x(1\rangle_p \right] \tag{17}$$

Factoring out the H

$$\underset{mxp}{Z} = \underset{mxp}{H}\,[x(l)\rangle_1, \ . \ . \ . \ , x(l)\rangle_p] = \underset{mxp}{H}\,\underset{pxp}{X} \tag{18}$$

If the known input vectors are linearly independent, that is the inverse matrix X^{-1} exists, then we can solve for H as

$$\underset{mxp}{Z}\,\underset{pxp}{X^{-1}} = \underset{mxp}{H} \tag{19}$$

Noise Conditions

The report covers the following cases in the respective order.

Case I. Scalar Case (scalar mean).

The noisy scalar case ($m = p = 1$) yields

$$z_k = \alpha_0 + v_k = \hat{a} + e_k \tag{20}$$

$$k = 1, 2, \ldots , k_{max}$$

where α_0 is the "true parameter" and $\hat{a}$ is our estimate of the parameter α_0 based on k_{max} observations. An unweighted and a weighted estimate will be derived. The error e_k is the observation minus our estimate $\hat{a}$ (the residuals).

Case II. Vector Mean Case.

The multivariable or vector case corresponds to

$$z(m)\rangle_k = \alpha(m)\rangle_0 + v(m)\rangle_k = \hat{a}(m)\rangle + e(m)\rangle_k \ . \tag{21}$$

Instead of one parameter in equation (20), we want to estimate m parameters in equation (21).

Case III. Scalar Polynomials.

The approximation of a function with a polynomial using unweighted and weighted least squares considers

$$z_k = \alpha_0 + \alpha_1 x_k + \alpha_2 x_k^2 + \ldots, \alpha_{p-1} x_k^{p-1} + v_k \tag{22}$$

$$= \hat{a}_0 + \hat{a}_1 x_k + \hat{a}_2 x_k^2 + \ldots, \hat{a}_{p-1} x_k^{p-1} + e_k \tag{23}$$

or in a vector-space setting

$$z_k = (\alpha_0, \alpha_1 \ldots, \alpha_{p-1}) \begin{pmatrix} 1 \\ x \\ x^2 \\ \cdot \\ \cdot \\ \cdot \\ x^{p-1} \end{pmatrix} + v_k \tag{24}$$

$$z_k = (\hat{a}_0, \hat{a}_1, \hat{a}_2 \ldots \hat{a}_{p-1}) \begin{pmatrix} 1 \\ x \\ x^2 \\ \vdots \\ x^{p-1} \end{pmatrix} + e_k \tag{25}$$

Define the p-dimensional parameter row vectors as

$$\langle p)\beta = (\beta_1, \beta_2, \ldots, \beta_p) = (\alpha_0, \alpha_1, \ldots \alpha_{p-1}) \tag{26}$$

and

$$\langle p)\hat{b} = (\hat{b}_1, \hat{b}_2, \ldots \hat{b}_p) = (\hat{a}_0, \hat{a}_1 \ldots \hat{a}_{p-1}) \tag{27}$$

and the p-dimensional column vector of data as

$$f(p\rangle_k = \begin{pmatrix} 1 \\ x \\ x^2 \\ \cdot \\ \cdot \\ \cdot \\ x^{p-1} \end{pmatrix} \tag{28}$$

Using the above relations

$$z_k = \langle\beta \; f\rangle_k + v_k = \langle\hat{b} \; f\rangle_k + e_k \tag{29}$$

If we now have an experiment with k observations (or a sample of size k) then the 2k scalar equations

$$z_1 = \langle\beta \; f\rangle_1 + v_1 = \langle\hat{b} \; f\rangle_1 + e_1 \tag{30}$$

$$\vdots$$

$$z_k = \langle\beta \; f\rangle_k + v_k = \langle\hat{b} \; f\rangle_k + e_k$$

can be written as two vector equations in k-space as

$$(z_1, z_2 \ldots z_k) = (\langle\beta \; f\rangle_1 , \langle\beta \; f\rangle_2 , \ldots \langle\beta \; f\rangle_k) \tag{31}$$

$$+ (v_1, v_2, \ldots v_k)$$

$$= (\langle\hat{b} \; f\rangle_1 , \langle\hat{b} \; f\rangle_2 , \ldots \langle\hat{b} \; f\rangle_k) \tag{32}$$

$$+ (e_1, e_2, \ldots, e_k)$$

Factoring out the vectors $\langle\beta$ and $\langle\hat{b}$

$$\langle z = \langle\beta \left[f\rangle_1 \; f\rangle_2 , \ldots f\rangle_k \right] + \langle k)v \tag{33}$$

$$= \langle\hat{b} \left[f\rangle_1 , \ldots f\rangle_k \right] + \langle k)e \tag{34}$$

Define the pxk data matrix as

$$\underset{pxk}{F} = \left[f(p\rangle_1 , f(p\rangle_2 , \ldots f(p\rangle_k \right] \tag{35}$$

$$= \begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_k \\ x_1^2 & x_1^2 & x_k^2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ x_1^{p-1} & x_2^{p-1} & x_k^{p-1} \end{bmatrix} \tag{36}$$

In vector matrix form equation (33) and (34) become

$$\langle k) z = \langle p)\beta \underset{pxk}{F} + \langle k) v = \langle \hat{b} \underset{pxk}{F} + \langle k) e \tag{37}$$

If we transpose to a column vector

$$z(k\rangle = \underset{kxp}{F^T}\, \beta(p\rangle + v(k\rangle = \underset{kxp}{F^T}\, \hat{b}\rangle + e(k\rangle \tag{38}$$

Note that the vector equation (38) looks like equation (7) except that m is replaced by k the sample-size which can become quite large, whereas m is equal to and generally less than p (since we can not instrument all variables of interest). We may also consider the matrix H as a mapping down to a sub-space whereas F is a mapping up or down depending on the size of k.

Case IV. Vector Polynomials

Approximating components of a vector with time polynomials, for example missile position vector, velocity vector etc., yields for n variables

$$z_1(k) = \beta_{11} + \beta_{21}\, x_k + \beta_{31}\, x_k^2 + \dots + v_{1k} \tag{39}$$

$$\vdots$$

$$z_n(k) = \beta_{1n} + \beta_{2n}\, x_k + \beta_{3n}\, x_k^2 + \dots + v_{nk}$$

or as inner products

$$z_1(k) = \underset{1}{\langle p)\beta}\, \underset{k}{f\rangle} + v_{1k} = \underset{1}{\langle \hat{b}}\, \underset{k}{f\rangle} + e_{1k} \tag{40}$$

$$\vdots$$

$$z_n(k) = \underset{n}{\langle p)\beta}\, \underset{k}{f\rangle} + v_n(k) = \underset{n}{\langle \hat{b}}\, \underset{k}{f\rangle} + e_{nk}$$

The kth observation of the n-dimensional vector is

$$z(k)(n\rangle = \begin{pmatrix} {}^{1}\langle p)\beta\ f\rangle_k \\ \cdot \\ \cdot \\ \cdot \\ {}^{n}\langle p)\beta\ f\rangle_k \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \\ \\ v_n \end{pmatrix} \tag{41}$$

$$= \begin{pmatrix} {}^{1}\langle\beta \\ \cdot \\ \cdot \\ \cdot \\ {}^{n}\langle p)\beta \end{pmatrix} f\rangle_k + v\rangle_k \tag{42}$$

$$z(n\rangle_k = \underset{nxp}{B}\ f(p\rangle_k + v(n\rangle_k \tag{43}$$

Forming a row of column vectors for k observations we obtain

$$\left[z(n\rangle_1 \ , \ z(n\rangle_2 \ , \ \ldots \ z(n\rangle_k \right] = \underset{nxk}{Z} \tag{44}$$

or

$$\underset{nxk}{Z} = \underset{nxp}{B}\ \underset{pxk}{F} + \underset{nxk}{V} = \hat{B}F + E \tag{45}$$

The next section develops the concepts of variance matrices around Case I, (the most simple case we can discuss) and applies the variance to weighted least squares.

The two age-old techniques of unweighted and weighted least squares are developed in a vector space setting.

SECTION II. ESTIMATION OF A CONSTANT SCALAR PLUS NOISE

Case I. Scalar Case

Consider the simple case of equation (I-20) where

$$z_k = \alpha_0 + v_k \tag{1}$$

where $k = 1, 2, \ldots, k_{max}$ is the number of observations.

Suppose we want to estimate α_0 based on a sequence of size k outputs, and designate our estimate of the parameter based on k values of z as $\hat{a}(k)$ or

$$z_k = \alpha_0 + v_k = \hat{a}(k) + \hat{e}_k, \tag{2}$$

The 2k equations in one space

$$\begin{aligned} z_1 &= \alpha_0 + v_1 = \hat{a}(k) + \hat{e}_1 \\ z_2 &= \alpha_0 + v_2 = \hat{a}(k) + \hat{e}_2 \\ &\vdots \\ z_k &= \alpha_0 + v_k = \hat{a}(k) + \hat{e}(k) \end{aligned} \tag{3}$$

can be written as two row-vector equations in k-space as

$$\begin{aligned} (z_1, z_2, \ldots, z_k) &= (\alpha_0, \alpha_0 \ldots \alpha_0) + (v_1, v_2 \ldots v_k) \\ &= (\hat{a}, \hat{a} \ldots \hat{a}) + (\hat{e}_1, \ldots, \hat{e}_k) \end{aligned} \tag{4}$$

we can factor α_0 and $\hat{a}$ out of the row vector and obtain

$$\begin{aligned} (z_1, z_2, \ldots z_k) &= \alpha_0(1, 1, \ldots 1) + (v_1, \ldots v_k) \\ &= \hat{a}(1, 1, 1, \ldots 1) + (\hat{e}_{11} \ldots \hat{e}_k) \end{aligned} \tag{5}$$

Define the <u>sum</u> - <u>vector</u> as

$$\langle k) 1 = (1, 1, 1 \; . \; . \; . \; 1) \tag{6}$$

hence

$$\langle k) z = \alpha_0 \langle 1 + \langle v = \hat{a} \langle 1 + \langle \hat{e} \tag{7}$$

Note that equation (7) is two vector equations in k-space.

<u>Unweighted Least Squares</u>

We obtain the unweighted least squares estimate simply by averaging all of the data, or all of the equations of (3), or

$$z_1 + z_2 + \; . \; . \; . \; + z_k = k \; \hat{a}(k) + \hat{e}_1 + \hat{e}_2 + \; . \; . \; . \; \hat{e}_k \tag{8}$$

and equating the sums of the $\hat{e}_k$'s to zero, that is

$$\hat{e}_1 + \hat{e}_2 + \; . \; . \; . \; + \hat{e}_k = 0 \tag{9}$$

The summation of the k scalar equations and averaging is equivalent to multiplying vector equation (7) by the column vector $1(k\rangle$ where

$\langle k) 1 \; 1 (k \rangle$

$$\langle k) 1 \; 1 (k \rangle = \underset{1xk}{(1, 1, \; . \; . \; . \; 1)} \; \underset{kx1}{\begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}} = k \qquad 10 \tag{10}$$

hence

$$\frac{\langle z \; 1 \rangle}{\langle 1 \; 1 \rangle} = \hat{a}(k) = \alpha_0 + \frac{\langle v \; 1 \rangle}{\langle 1 \; 1 \rangle} \tag{11}$$

Clearly the relation of equation (9) is equivalent to orthogonality since

$$\langle \hat{e}\ 1 \rangle = \hat{e}_1 + \ . \ . \ . + \hat{e}_k = 0 \tag{12}$$

as shown in Figure (1)

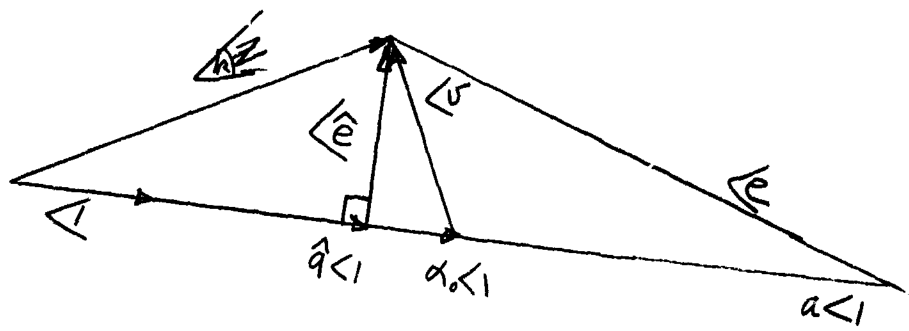

Fig. (1) Vector in k-Space

The hat symbol on the $\hat{a}$ corresponds to the value of a (any real number) which makes the residual vector $\langle e$ perpendicular to the sum vector $\langle 1$. This minimum magnitude vector is designated by $\langle \hat{e}$ and satisfies equation (12). See reference (4) in which the least squares relations are derived via gradient methods using partial derivatives and via completely algebraic methods using orthogonal projections.

Observe that the noise sums are not zero

$$v_1 + v_2 + \ . \ . \ . + v_k = \langle v\ 1 \rangle \neq 0 \tag{13}$$

in general since we can not control the true noise values.

Note that in scalar summation form equations (11) and (12) are

$$\hat{a}(k) = \frac{1}{k_{max}} \sum_{k=1}^{k_{max}} z_k = \alpha_0 + \sum_{k=1}^{k_{max}} v_k \frac{1}{k_{max}} \tag{14}$$

and

$$\sum_{k=1}^{k_{max}} e_k = 0 \tag{15}$$

Thus far we have made only one statement (equation (13)) about the statistical characteristics of the noise v_k.

The error in the estimate of the parameter is by equation (11)

$$a_0 - \hat{a}(k) = -\frac{\langle v\, 1\rangle}{k} = \tilde{a}(k) \qquad (16)$$

and the square of the error in the parameter estimation is

$$\tilde{a}^2(k) = \frac{\langle k)1\; v(k\rangle\langle 1)v\; 1(k\rangle}{k^2} \qquad (17)$$

where the kxk square matrix (dyad) is

$$v(k\rangle\langle k)v = \begin{pmatrix} v^1 \\ v^2 \\ \cdot \\ \cdot \\ \cdot \\ v^k \end{pmatrix} (v_1, v_2, \ldots v_k) = \begin{pmatrix} v^1 v_1 & \ldots & v^1 v_k \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ v^k v_1 & & v^k v_k \end{pmatrix} \qquad (18)$$

In summation form equation (16) is

$$\tilde{a}(k) = -\sum_{k=1}^{k_{max}} v_k \frac{1}{k_{max}} \qquad (19)$$

and the square of the error equation (17) is

$$\tilde{a}^2(k) = \left(\sum_{k=1}^{k_{max}} u_k\right)^2 \frac{1}{k^2_{max}} \qquad (20)$$

The above two equations and the scalar summation equations of (14) and (15) require only a knowledge of real variables or real field algebra and summation index "rules". The representations of equation (11) and (17) require a knowledge of vector inner-products and "outer" or dyadic products, where the dyad of equation (18) is a rank-one kxk matrix. From the above we note some of the simple but basic differences between the state-space approach versus the older say-it-with-summation-sigma-signs.

Summarizing, the unweighted estimate by equation (11) is

$$\hat{a}(k) = \langle k)z\ 1(k\rangle \frac{1}{k} \tag{21}$$

and the square of the <u>error in the estimate of the parameter</u> is by equation (17)

$$\tilde{a}^2(k) = \langle 1[v(k\rangle\langle k)v]\ 1(k\rangle \frac{1}{k^2} \tag{22}$$

Note that we can consider the arithmetic mean (unweighted) case as an equi-weight case where each data-point is weighted by 1/k or as a sequence, or vector, of weights

$$\frac{1}{k}\langle 1 = (1/k,\ 1/k,\ .\ .\ .\ 1/k) \tag{23}$$

We may now ask the question: Can we obtain an estimate of α_0 which is "better" than equation (21) and which has a smaller numerical value of error-square of equation (22)?

The next section will derive a sequence of weights such that a weighted estimate of the parameter is a linear combination of the weights and the data, that is

$$\hat{a}_w = z_1\, w_1 + z_2\, w_2 + \ .\ .\ .\ + z_k\, w_k \tag{24}$$

In a vector-space setting, we seek to find a column vector of weights $w\rangle$ such that

$$\hat{a}_w = \langle k)z\ w(k\rangle\ , \tag{25}$$

and that on the average, equation (24) is "better in some sense" than equation (21).

WEIGHTED LEAST SQUARES

The application of weighted least-squares and the derivation of the equations are developed in this section for the scalar case. The application to the observational data in the context of this report is equivalent to a statistical calibration of the instrument (that is a calibration with respect to its noise characteristics).

Noise Considerations and Noise Variance Matrix.

Before we utilize the instrument for experiments or tests we can calibrate the noise by setting $x(1)$ (the input) equal to zero, hence the only output is v_k. Many experiments exist in which we cannot control the input, for example set the input equal to zero, in order to calibrate the instrument. An example is a missile flight test for which we want to calibrate a tracking radar with respect to its noise for that region of tracking space. In this case one needs a higher quality trajectory measuring device (optical perhaps) or else a minimum of three redundant sensors such that differencing makes the calibration results independent of the trajectory (see reference (5)). The remainder of the discussions in this report assumes we can control the inputs to zero.

Many instrumentation systems observing dynamical processes have an upper bound on the observation time, which in conjunction with samples per second sets a maximum sample size, say k_{max}. If we now have time in advance to prepare for the test, to study the outputs for samples up to k_{max}, say

$$(v_1, v_2, \dots v_{kmax}) = \langle k)v \qquad (26)$$

and repeat the sequence (reset the instrument) j_{max} vectors each of dimension k_{max}. That is

$$\langle k)_j v = (v_1, v_2, \dots v_{k_{max}})_j \qquad (27)$$

where $j = 1 \dots j_{max}$ where j_{max} may be whatever economical number we can afford. We certainly can not calibrate to infinity.

The k-discrete points may be taken as points off of a continuous curve $v_j(t)$ as shown in Figure (2)

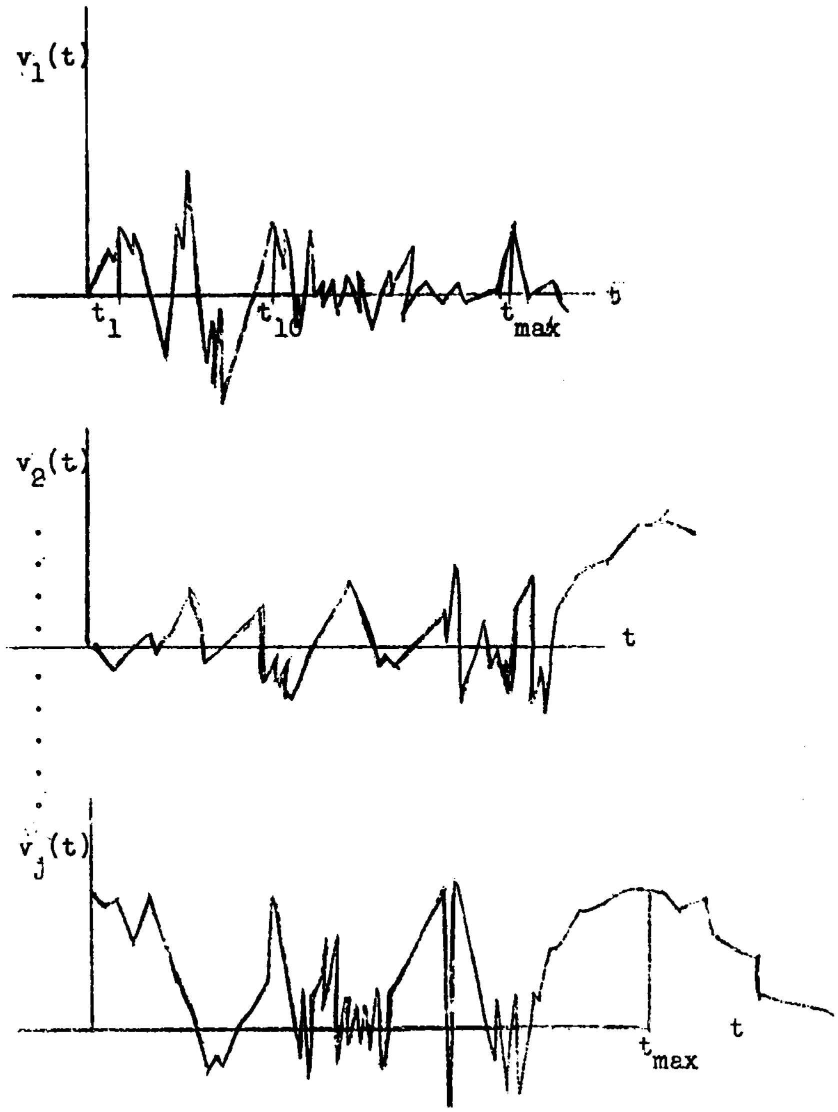

Fig (2) Sequences of Time-Correlated Noise

For example, suppose we are planning to use the instrument in a number of tests or experiments such that this particular device is to measure a constant during each test. The duration of each test is such that this particular instrument takes k_{max} samples. The k_{max} is usually dictated by economy of data processing, time-sharing of a complete system of sensor outputs via telemetry, etc.

We can record the j_{max} sequences (row vectors) each of dimension k_{max} or sequentially feed the data output into a digital computer data-processing program.

What should we compute in the program? Let us return briefly to the unweighted case where the unweighted estimate by equation (11) is

$$\hat{a}(k) = \frac{\langle k)\, z\, 1(k\rangle}{\langle 1\, 1(k\rangle} \tag{28}$$

and the error square term is

$$\tilde{a}^2(k) = \left\langle 1 \left[v(k\rangle\langle k)\, v \right] 1 \right\rangle \frac{1}{k^2} \quad . \tag{29}$$

In an actual test with an input different from zero we do not know the values of $(v_1, v_2, \dots v_k)$, hence we can not compute $\tilde{a}^2(k)$. For example, suppose some arbitrary noise sequence $\langle \underset{j}{v}$ occurs during the test, then the parameter estimation error based on a sample of size k occuring as a result of the jth noise sequence is

$$\tilde{a}_j^2(k) = \left\langle 1 \left[\underset{j}{v}\rangle\langle \overset{j}{v} \right] 1 \right\rangle \frac{1}{k^2} \tag{30}$$

The average error over all j_{max} noise sequences is

$$\sigma_{\tilde{a}\tilde{a}}(k) = \frac{1}{j_{max}} \left[\tilde{a}_1^2(k) + \tilde{a}_2^2(k) + \dots + \tilde{a}_j^2(k) + \dots + \tilde{a}_{j_{max}}^2(k) \right] \tag{31}$$

or in summation form

$$\sigma_{\tilde{a}\tilde{a}}(k) = \left(\sum_{j=1}^{j_{max}} \tilde{a}_j^2(k) \right) \frac{1}{j_{max}} \quad . \tag{32}$$

The scalar $\sigma_{\tilde{a}\tilde{a}}(k)$ is called the <u>variance of the estimate of the parameter</u>, or the average error in the estimate of the parameter over all experiments j.

If we use the dyad expression of equation (17) in equation (31) we obtain

$$\sigma_{\tilde{a}\tilde{a}}(k) = \frac{1}{j_{max}} \left[\langle 1\, \underset{1}{v}\rangle\langle \overset{1}{v}\, 1\rangle + \langle 1\, \underset{2}{v}\rangle\langle \overset{2}{v}\, 1\rangle + \dots + \langle 1\, \underset{j_{max}}{v}\rangle\langle \overset{j_{max}}{v}\, 1\rangle \right] \frac{1}{k^2} \tag{33}$$

Factoring out the summation vector from each end

$$\sigma_{\tilde{a}\tilde{a}}(k) = \frac{1}{k^2}\left\langle 1 \left[\frac{v\rangle_1 \langle^1 v + \ldots + v\rangle_j \langle^j v + \ldots + v\rangle_k \langle^k v}{j_{max}} \right] 1 \right\rangle \tag{34}$$

or in summation form

$$\sigma_{aa} = \frac{1}{k^2}\left\langle 1 \left[\frac{1}{j_{max}} \sum_{j=1}^{j_{max}} v\rangle_j \langle^j v \right] 1 \right\rangle \tag{35}$$

The kxk matrix is the arithmetic mean of the j_{max} dyads and will be designated as

$$\underset{kxk}{\Sigma_{vv}} \equiv \sum_{j=1}^{j_{max}} v(k\rangle_j \langle^j k)v \;\; \frac{1}{j_{max}} \tag{36}$$

We shall also occassionally use the notation

$$\underset{kxk}{\Sigma_{vv}} = \underset{kxk}{Q(k)} \tag{37}$$

as occurs in many of the modern estimation publications.

We shall also use the notation or symbol for the "expectation operator"

$$E_j \left\{ v\rangle_j \langle^j v \right\} = \lim_{j_{max}\to\infty} \left[\sum_{j=1}^{j_{max}} v\rangle_j \langle^j v \right] \frac{1}{j_{max}} \tag{38}$$

However, from the practical world standpoint we assume

$$\lim_{j_{max}\to\infty} \sum_{j=1}^{j_{max}} v\rangle_j \langle^j v = \sum_{j=1}^{j_{max}} v\rangle_j \langle^j v \;\; \frac{1}{j_{max}} + E_r \tag{39}$$

where the error matrix E_r is almost zero and j_{max} is dictated by a large enough finite-population to be statistically representative of the infinite population and economically available.

Hence, throughout the paper we assume

$$E_j \left\{ \underset{j}{v\rangle} \overset{j}{\langle v} \right\} = \sum_{j=1}^{j_{max}} \underset{j}{v\rangle} \overset{j}{\langle v} \frac{1}{j_{max}} \tag{40}$$

or the <u>expectation-operator</u>, as applied, is merely the average of the dyad sums.

During an actual experiment $\underset{j}{\langle v}$ comes from an infinite universe or population; but from the real-world calibration standpoint we must make computations based on a countable finite and economical population.

Note that R is not the variance with respect to the noise mean; however we shall hence-forth refer to it as the instrument or merely <u>noise variance matrix</u>.

It is the variance with respect to a different "origin" not the mean as origin.

The variance of the noise with respect to its mean is

$$E_j \left\{ \left(\underset{j}{v\rangle} - \hat{v}\rangle \right) \left(\overset{j}{\langle v} - \langle \hat{v} \right) \right\} \tag{41}$$

where the mean is

$$\underset{j}{\hat{v}\rangle} = \left(\sum_{j=1}^{j_{max}} \underset{j}{v\rangle} \right) \frac{1}{j_{max}} \tag{42}$$

and can be computed to give us more information about the noise characteristics.

The expression of equation (41) is the most familiar expression for a variance matrix.

A recursive method for digitally computing the matrix Q of equation (40) for any number of vectors $\underset{j}{\langle v}$ is given in appendix

The expected error in the estimate (one-dimensional ellipsoid of uncertainty) of the parameter by equation (22) and equation (36) for unweighted estimation is

$$\alpha_{\hat{a}\hat{a}}(k) = \frac{1}{k} \langle k) 1 \Sigma_{vv} 1 (k \rangle \tag{43}$$

<u>Derivation of the Weights.</u>

Consider the data-vector $\langle k)z$ of equation (5) which occurs as the outcome of an experiment "confused" by an arbitrary noise sequence $\underset{j}{\langle k)}v$, then

$$\underset{j}{\langle z)} = \alpha_0 \langle 1 + \underset{j}{\langle v} = \hat{a}_j(k) \langle 1 + \underset{j}{\langle e} \tag{44}$$

Note that the parameter α_0 does not change with j (that is the exciting noise sequence $\underset{j}{\langle v}$) but all variables subscripted with j do.

We may also take the state of mind that equation (44) is the result of repeating the experiment j time and $\underset{j}{\langle z}$ is the data sequence occuring as a result of α_0 and $\underset{j}{\langle v}$.

We now seek a sequence of k scalar weights designated as a column vector (independent of j)

$$w(k\rangle = \begin{pmatrix} w^1 \\ w^2 \\ \cdot \\ \cdot \\ \cdot \\ w^k \end{pmatrix} \tag{45}$$

such that the inner-product of $w\rangle$ with equation (44) is

$$\langle \underset{j}{z}\, w \rangle = \alpha_0 \langle 1\, w \rangle + \langle \underset{j}{v}\, w \rangle = a_j(k) \langle 1\, w \rangle + \langle \underset{j}{e}\, w \rangle \tag{46}$$

where the conditions hold

$$\langle 1\, w \rangle = 1 \tag{47}$$

$$\langle \underset{j}{e}\, w \rangle = 0 \tag{48}$$

Note that we want a single vector $w\rangle$ to be used for all possible noise sequences $\underset{j}{v}\rangle$

Using the constraints of equation (47), (48) in equation (46)

$$\langle \underset{j}{z}\; w\rangle = \alpha_0 + \langle \underset{j}{v}\; w\rangle = \hat{a}_{jw}(k) \tag{49}$$

or the weighted estimate of the single parameter based on k samples is a linear-conbination of the data

$$\hat{a}_{jw}(k) = \langle \underset{j}{z}\; w\rangle = z_{1_j}\; w_1 + \ldots + z_{k_j}\; w_k. \tag{50}$$

The error in the estimate of the parameter is

$$\alpha_o - \hat{a}_{j_w}(k) \equiv \tilde{a}_{j_w}(k) = -\langle \underset{j}{v}\; w\rangle . \tag{51}$$

Since the inner-product of two vectors is a scalar and commutativity holds

$$\tilde{a}_{j_w}(k) = -\langle w\; \underset{j}{v}\rangle . \tag{52}$$

The square of the error in the weighted estimate of the parameter by equation (51) and equation (52) is

$$(\tilde{a}_{j_w}(k))^2 = \langle w\; \underset{j}{v}\rangle\langle \overset{j}{v}\; w\rangle \tag{53}$$

or

$$(\tilde{a}_{j_w}(k))^2 = \langle w[\underset{j}{v}\rangle\langle \overset{j}{v}]w\rangle \tag{54}$$

The average value of the error-squared for all possible noise sequences j is

$$[\tilde{a}^2_{1_w}(k) + \tilde{a}^2_{2_w}(k) + \ldots + \tilde{a}^2_{j_w}(k)]\; \frac{1}{j_{max}}$$

$$= \sum_{j=1}^{j_{max}} [\tilde{a}_{j_w}(k)]^2\; \frac{1}{j_{max}} = \sigma_{\tilde{a}\tilde{a}_w}(k) \tag{55}$$

which is the <u>weighted variance of the estimate of the parameter</u>.

As before the "expected" error square is

$$E_j\{\tilde{a}^2_{j_w}(k)\} \approx \sum_{j=1}^{j_{max}} \tilde{a}^2_{j_w}(k)\, \frac{1}{j_{max}} \, . \tag{56}$$

If we now use the dyadic-expression of equation (52) in equation (55) we obtain

$$\sigma_{\tilde{a}\tilde{a}_w} = \frac{\langle w \left\{ v_1 \times v^1 + \dots + v_{j_{max}} \times v^{j_{max}} \right\} w \rangle}{j_{max}} \tag{57}$$

or by equation (36)

$$\sigma_{\tilde{a}\tilde{a}_w} = \langle w \, \Sigma_{vv} \, w \rangle \tag{58}$$

Equation (58) is quadratic in the unknown vector $\langle w$. We now seek a vector $\langle w$ which minimizes the variance of the estimate of the parameter over all experiments (or noise sequences j) and also satisfies the constraint of equation (47).

The solution, by appendix C, equation (15) is

$$\langle w = \frac{\langle 1 \, \Sigma^{-1}_{vv}}{\langle 1 \, \Sigma^{-1}_{vu} \, 1 \rangle} \tag{59}$$

Utilizing $\langle w$ in equation (58)

$$\sigma_{\tilde{a}\tilde{a}_w} = \frac{1}{\langle 1 \, \Sigma^{-1}_{uu} \, 1 \rangle} \tag{60}$$

and the weighted estimate by equation (59) in equation (49) is

$$\hat{a}_w(k_{max}) = \frac{\langle k) z \, \Sigma^{-1}_{uu} \, 1(k \rangle}{\langle k) 1 \, \Sigma^{-1}_{uu} \, 1(k \rangle} \, . \tag{61}$$

SECTION III

ESTIMATION OF A CONSTANT VECTOR PLUS NOISE

This section develops the unweighted and weighted estimation equations for a constant vector plus noise. Utilizing the concepts and notation for the scalar case except now we assume that there are m measurement variables (z_1, z_2, . . . z_m), and an experiment or test for which we take k_{max} observations. During the test there will be some noise vector sequence V(j)

$$\left[v(1)(m\rangle, v(2)(m\rangle, \ldots v(k\rangle, v(k_{max}\rangle \right]_j = \underset{mxk_{max}}{V(j)} \qquad (1)$$

out of a possible j_{max} sequences

$$\left[\underset{mxk_{max}}{V(1)}, \ldots \underset{mxk_{max}}{V(j)}, \ldots \underset{mxk_{max}}{V(j_{max})} \right] = \langle j_{max})V \qquad (2)$$

where j_{max} is infinite, and $\langle j_{max})V$ designates a "row vector or matrix" of mxk matrices.

We designate the kth observation and its relation to the noise as

$$z(k)(m\rangle_j = \alpha(m\rangle + v(k)(m\rangle_j \qquad (3)$$

where the unknown constant vector is $\alpha(m\rangle$. One may interpret the constant vector of equation (3) and equation (1-8) hence

$$\alpha(m\rangle = \underset{mxp}{H_0}\, x(1)(p\rangle \qquad (4)$$

If we form a data-matrix by a row of column vectors

$$\left[z(m\rangle_1, z(m\rangle_2, \ldots z(m\rangle_k \right] = \underset{mxk}{Z} = \left[\alpha\rangle, \alpha\rangle, \ldots \alpha\rangle \right]$$

$$+ \left[v\rangle_1 \ldots v\rangle_k \right] \qquad (5)$$

and factor out the $\alpha\rangle$

$$\underset{m\times k}{Z} = \alpha(m\rangle\langle k)1 + \underset{m\times k}{V}. \tag{6}$$

Consider $a(m\rangle_j$ an arbitrary m dimensional vector and the error or residual vector $e\rangle_j$ such

$$z(k)(m\rangle_j = a(m\rangle_j + e(k)(m\rangle_j = \alpha\rangle + v(k)(m\rangle_j. \tag{7}$$

The data-matrix equations for all k observations become

$$\underset{m\times k}{Z(j)} = \alpha(m\rangle\langle k)1 + \underset{m\times k}{V(j)} = a\rangle_j\langle 1 + E(j) \tag{8}$$

If we subtract the terms

$$\left[\, a\rangle_j - \alpha\rangle \right]\langle k) = E(j) - V(j). \tag{9}$$

Unweighted Least Squares Estimate

The arithmetic average of the vectors using none of the noise characteristics yields

$$\frac{z(m\rangle_{1j} + z\rangle_{2j} + \ .\ .\ .\ + z\rangle_{jk_{max}}}{k_{max}} = \hat{a}(m\rangle_j \tag{10}$$

which is the unweighted least-squares estimate.

In vector-matrix form we obtain equation (10) by multiplying equation (9) by the column vector

$$1\rangle \frac{1}{k}$$

$$\frac{Z_j 1\rangle}{\langle 11\rangle} = \hat{a}_j\rangle = \alpha\rangle + \frac{V_j 1\rangle}{k} \qquad (11)$$

with the constraint of

$$E_j 1\rangle = 0\rangle = e_1\rangle + \ldots + e_k\rangle \qquad (12)$$

The error in the unweighted estimate resulting from the jth noise sequence by equation (11) is

$$\alpha\rangle - \hat{a}_j\rangle = \tilde{a}(m_j\rangle = -\underset{mxk}{V_j} 1(k\rangle \frac{1}{k_{max}} \qquad (13)$$

Transposing (13)

$$\langle m_j)\tilde{a} = -\langle k)1 \, \underset{kxm}{V_j^T} \, \frac{1}{k_{max}} \qquad (14)$$

The dyadic product of (14) and (13) is the mxm matrix

$$\tilde{a}(m_j\rangle\langle m_j)\tilde{a} = \underset{mxk}{V_j} \, 1(k\rangle\langle k)1 \, \underset{kxm}{V^T} \, \frac{1}{k^2_{max}} \qquad (15)$$

The variance matrix of the unweighted estimate of the parameters is the average over all noise sequences j and is the symmetric matrix

$$\underset{mxm}{\Sigma_{\tilde{a}\tilde{a}}} = E_j \left\{ \tilde{a}(m_j\rangle\langle m_j)\tilde{a} \right\} = \sum_{j=1}^{j_{max}} a_j\rangle\langle a_j \frac{1}{j_{max}} \qquad (16)$$

The above mxm matrix represents the uncertainty ellipsoid in m-space. The trace of the dyad of equation (15) is the inner-product term

$$tr\left[a_j\rangle\langle a_j \right] = \langle a_j \, a_j\rangle = \langle 1 \, V_j^T \, V_j \, 1\rangle \frac{1}{k^2_{max}} \qquad (17)$$

and the trace of equation (16) is

$$\mathrm{tr}\,\Sigma_{\tilde{a}\tilde{a}} = E_j \left[\langle \tilde{a}_j \; \tilde{a}_j \rangle \right]$$

$$= \frac{1}{k^2_{max}} \left[\langle 1 \; \frac{V_1^T V + V_2^T V_2 + \ldots + V^T_{j_{max}} V_{j_{max}}}{j_{max}} \right] 1\rangle \tag{18}$$

$$\mathrm{tr}\,\Sigma_{\tilde{a}\tilde{a}} = \langle 1 \underset{kxk}{Q_{VV}} 1\rangle \frac{1}{k^2_{max}} \tag{19}$$

where Q is the average of the matrix products

$$\underset{kxk}{Q_{VV}} = \sum_{j=1}^{j_{max}} \underset{kxm}{V(j)^T} \underset{mxk}{V(j)} \frac{1}{j_{max}} \tag{20}$$

Weighted Least Squares Estimate

We now seek an estimate with a smaller ellipsoid of uncertainty. Consider equation (6)

$$\underset{mxk}{Z_j} = \alpha\rangle\langle 1 + V_j = \tilde{a}_j\rangle\langle 1 + E_j \tag{21}$$

We need a k_{max} dimensional column vector $w\rangle$ such that

$$Z_j w\rangle = \alpha\rangle\langle 1 \; w\rangle + V_j w\rangle = \tilde{a}_j\rangle\langle 1 \; w\rangle + E_j w\rangle \tag{22}$$

satisfying the conditions

$$\langle 1 \; w\rangle = 1 \tag{23}$$

$$E_j w\rangle = 0(m\rangle \tag{24}$$

then

$$Z_j w\rangle = \hat{a}\rangle_{jw} \tag{25}$$

Using the constraint equations (23) and (24) equation (22) becomes

$$\underset{m\times k}{Z_j}\, w(k)\rangle = \alpha\rangle + V_j w\rangle = \hat{a}(m)\rangle_{jw} \tag{26}$$

Note that the weighted estimate is a linear combination of the observation vector

$$Zw(k)\rangle = z_1\rangle\, w_1 + z_2\rangle\, w_2 + \dots + z_{k_{max}}\rangle\, w_k = \hat{a}\rangle \tag{27}$$

The error in the estimate by equation (26) is

$$\alpha\rangle - \hat{a}\rangle_{wj} = \tilde{a}\rangle_{jw} = -\,V_j w\rangle \tag{28}$$

and transposing equation (28)

$$\tilde{a}\rangle_{jw} = \langle w\, V_j^T\,. \tag{29}$$

The mxm random matrix dyadic product is

$$\tilde{a}\rangle_{jw}\langle\tilde{a}^{jw} = V_j w\rangle\langle w V_j^T \tag{30}$$

The weighted variance of the estimate is the symmetric mxm matrix

$$E_j\left\{\tilde{a}\rangle_{jw}\langle\tilde{a}^{jw}\right\} = \underset{mxm}{\not\angle}\,\tilde{a}\tilde{a}_w\,. \tag{31}$$

The trace of (30) is

$$\langle \tilde{a}_{jw} \tilde{a}_{jw} \rangle = \langle w V_j^T V_j w \rangle \tag{32}$$

and the trace of equation (31) is

$$\mathrm{tr}\, \Sigma_{\tilde{a}\tilde{a}_w} = E_j \left\{ \langle \tilde{a}_j \tilde{a}_{jw} \rangle \right\} = E_j \left[\langle w V_j^T V_j w \rangle \right] \tag{33}$$

$$= \left\langle w \left[\sum_{j=1}^{j_{max}} V_j^T V_j \frac{1}{j_{max}} \right] w \right\rangle \tag{34}$$

$$\mathrm{tr}\, \Sigma_{\tilde{a}\tilde{a}_w} = \langle w \underset{k\times k}{Q_{vv}} w \rangle \tag{35}$$

where Q_{vv} is given by equation (29)

The trace of the ellipsoid of equation (34) and the hyperplane constraint of equation (23) are exactly the same as the minimization problem of equation (II-58)and by equation(II-59) the weight vector is

$$w(k\rangle = \frac{\Sigma_{vv}^{-1}\, 1(k\rangle}{\langle 1 \Sigma_{vv}^{-1} 1 \rangle} \tag{36}$$

and the weighted estimate of the parameter vector is

$$\hat{a}(m\rangle_w = \underset{m\times k}{Z}\, w(k\rangle = \frac{Z_{m\times k} \Sigma_{vv}^{-1} 1\rangle}{\langle 1 \Sigma_{vv}^{-1} 1 \rangle} \tag{37}$$

with an ellipsoid of uncertainty by equation (36) in (30)

$$E\left[\tilde{a}_j \rangle \langle \tilde{a}^j \right] = E_j \frac{\left[V_j \Sigma_{vv}^{-1} 1 \rangle \langle 1 \Sigma_{vv}^{-1} V_j^T \right]}{\langle 1 \left(\Sigma_{vv}^{-1} \right) 1 \rangle^2} \tag{38}$$

SECTION IV.

POLYNOMIAL PARAMETER ESTIMATION

The classical approximation of a function by a pth degree polynomial and the weighted and unweighted least squares estimates of the parameter is developed.

Consider the polynomial

$$z_k = \alpha_0 + \alpha_1 x_k + \alpha_2 x_k^2 + \ldots + \alpha_{p-1} x_k^{p-1} + v_k \tag{1}$$

$$= a_0 + a_1 x_k + a_2 x_k^2 + \ldots a_{p-1} x_k^{p-1} + e_k \tag{2}$$

Separating the parameters

$$z_k = (\alpha_0, \alpha_1, \alpha_2, \ldots \alpha_{p-1}) \begin{pmatrix} 1 \\ x \\ x^2 \\ \cdot \\ \cdot \\ \cdot \\ x^{p-1} \end{pmatrix}_k + v_k \tag{3}$$

$$= (a_0, a_1, \ldots a_{p-1}) \begin{pmatrix} 1 \\ x \\ x^2 \\ \cdot \\ \cdot \\ \cdot \\ x^{p-1} \end{pmatrix} + e_k \tag{4}$$

or

$$z_k = \langle \alpha\, f \rangle_k + v_k = \langle a\, f \rangle_k + e_k \tag{5}$$

where

$$f(p\rangle_k = \begin{pmatrix} 1 \\ x \\ x^2 \\ \vdots \\ x^{p-1} \end{pmatrix}_k \tag{6}$$

If we have k_{max} observations packaged as a row vector we have

$$(z_1, z_2, \ldots z_{k_{max}}) = \langle k)z \tag{7}$$

$$= [\langle \alpha\, f\rangle_1, \langle \alpha\, f\rangle_2, \ldots, \langle \alpha\, f\rangle_{k_{max}}]$$

$$+ (v_1, v_2, \ldots, v_{k_{max}}) \tag{8}$$

$$= [\langle a\, f\rangle_1, \langle a\, f\rangle_2, \ldots, \langle a\, f\rangle_{k_{max}}]$$

$$+ [e_1, e_2, \ldots e_{k_{max}}] \tag{9}$$

Factoring out the row vector of parameters

$$\langle k)z = \langle p)\alpha \underset{pxk}{F} + \langle v \tag{10}$$

$$\langle z = \langle a\, F + \langle e, \text{ where} \tag{11}$$

$$\underset{pxk_{max}}{F} = [f\rangle_1, \ldots f\rangle_{k_{max}}] \tag{12}$$

Subtracting equation (11) from (10)

$$[\langle \alpha - \langle a]\, F = \langle e - \langle v \tag{13}$$

where we define the error in the parameters

$$\langle \tilde{a} = \langle \alpha - \langle a \tag{14}$$

<u>Unweighted Parameter Estimation</u>

This section obtains the unweighted estimate of the parameters and the variance of the estimate.

If we multiply equation (10) and (11) by the psuedo-inverse matrix F^+ which is a kxp matrix and

$$\underset{pxp}{F^+} = \underset{kxp}{F^T} \underset{pxp}{(FF^T)^{-1}} \tag{15}$$

with the one-sided inverse property

$$FF^T = I_{pxp} \tag{16}$$

then

$$\langle zF^T (FF^T)^{-1} = \langle \hat{a} \tag{17}$$

$$= \langle \alpha + \langle vF^T (FF^T)^{-1} \tag{18}$$

and

$$\langle \hat{e}F^T = \langle p)0. \tag{19}$$

Note that $\langle \hat{a}$ and $\langle \hat{e}$ correspond to those values of $\langle a$ and $\langle e$ such that $\langle \hat{e}\ \hat{e}\rangle$ is a minimum. The geometry and derivations are derived in reference (4) via partial derivatives and via orthogonal projections.

Differencing equation (17) and (18)

$$\langle \alpha - \langle \hat{a} = \langle \tilde{a}_j = - \langle v_j F^T(FF^T)^{-1} \tag{20}$$

where the j as before refers to the jth noise sequence. Transposing

$$\tilde{a}_j\rangle = (FF^T)^{-1} F v_j\rangle \tag{21}$$

The dyadic product of (20) and (21) is

$$\tilde{a}_j\rangle \langle^{j}\tilde{a} = (FF^T)^{-1}F\, v_j\rangle \langle^{j}vF^T(FF^T)^{-1} \tag{22}$$

and expected value over all noise sequences is

$$E_j \left\{ \tilde{a}_j\rangle \langle^{j}\tilde{a} \right\} = \Sigma_{\tilde{a}\tilde{a}\ pxp} \tag{23}$$

$$\not{\Sigma}_{\tilde{a}\tilde{a}} = (FF^T)^{-1} F \, E_j \left\{ v\rangle_j \, {}^j\langle v \right\} F^T (FF^T)^{-1} \tag{24}$$

where the noise characteristics are

$$\underset{k \times k}{Q} = E_j \left\{ v(k\rangle_j \, {}^j\langle k) v \right\} \tag{25}$$

Using equation (25) in equation (24) we see that the ellipsoid of uncertainty in p-space (the pxp symmetric matrix describing the variance of the estimate of the parameters) is

$$\underset{p \times p}{\not{\Sigma}_{\tilde{a}\tilde{a}}} = \underset{p \times p}{(FF^T)^{-1}} \underset{p \times k}{F} \underset{k \times k}{Q_{vv}} F^T (FF^T)^{-1}. \tag{26}$$

Weighted Least Squares

This section derives the classical weighted least-squares equations in a vector-space setting.

We seek a kxp matrix W such that post-multiplying equation (10) and (11)

$$\langle zW = \langle aFW + \langle vW \tag{27}$$

$$= \langle aFW + \langle eW \tag{28}$$

If the conditions of

$$\underset{(p \times k)}{F} \; \underset{(k \times p)}{W} = \underset{p \times p}{I} \tag{29}$$

and

$$\langle k)e \underset{k \times p}{W} = \langle p)e \tag{30}$$

then

$$\langle zW = \langle p)\hat{a}_W = \langle a + \langle vW . \tag{31}$$

TEXT NOT REPRODUCIBLE

If we factor W into its row-space, that is k vectors of dimension p

$$W = \begin{bmatrix} \overset{1}{\langle p)}w \\ \overset{2}{\langle p)}w \\ \vdots \\ \overset{k}{\langle p)}w \end{bmatrix} \tag{32}$$

then we can consider the p-dimensional row vector of parameters $\langle \hat{a}w$ as a linear combination of the scalar data and the weighting vectors

$$\underset{w}{\langle \hat{a}} = \langle zW = (z_1, z_2, \ldots z_k) \begin{bmatrix} \overset{1}{\langle p)}w \\ \vdots \\ \overset{k}{\langle w} \end{bmatrix} \tag{33}$$

or

$$\langle p)\hat{a}_w = z_1 \overset{1}{\langle p)}w + z_2 \overset{2}{\langle p)}w + \ldots + z_k \overset{k}{\langle p)}w \tag{34}$$

The error vector in the estimate of the parameters by equation (31) is

$$\langle \tilde{a}_{jw} = \langle \alpha - \underset{j}{\langle \hat{a}_w} = - \underset{j}{\langle v} W \tag{35}$$

where as before the j denotes the estimate resulting from the jth noise sequence

$$\underset{j}{\langle k_{max})}v = (v_1, v_2, \ldots v_{kmax})_j \tag{36}$$

and we want a W to be used for any of the j's, that is W is not a function of j.

The transpose of (35) is

$$\underset{wj}{\tilde{a}\rangle} = -W^T \underset{j}{v\rangle} \tag{37}$$

The outer-product and inner products respectively are

$$\tilde{a}\rangle_{wj}\,{}^{j}\langle\hat{a}_w = W^{T} y\rangle_{j}\,{}^{j}\langle y W \tag{38}$$

and

$$\mathrm{tr}\;\hat{a}_w\rangle_{j}\,{}^{j}\langle\hat{a}_w = \langle\hat{a}_j\;\hat{a}_j\rangle = \langle y_j W\; W^{T} y_j\rangle \tag{39}$$

By equation (Appendix B-79)

$$\frac{\partial \langle \hat{a}_{wj}\;\hat{a}_{w}\rangle}{\partial \underset{pxk}{W}} = W^{T} y\rangle_{j}\,{}^{j}\langle y\; 2 \tag{40}$$

Form the difference matrix

$$\tilde{a}(p\rangle_{j}\,{}^{j}\langle p)\tilde{a} - F\,W = \underset{pxp}{\Psi_{j}} \tag{41}$$

The sums over all j_{max} divided by j_{max} is

$$\frac{1}{j_{max}} \sum_{j=1}^{j_{max}} \Psi_j = E_j\left\{\tilde{a}\rangle_{j}\,{}^{j}\langle\hat{a}_w\right\} - FW \tag{42}$$

$$\Psi = \underset{pxp}{\Sigma_{\tilde{a}\hat{a}}} - \underset{pxp}{FW} \tag{43}$$

The trace of equation (43) is

$$\mathrm{tr}\;\Psi = \mathrm{tr}\,\Sigma_{\tilde{a}\tilde{a}} - \mathrm{tr}(FW) \tag{44}$$

The trace of equation (41)

$$\mathrm{tr}\ \Psi_j = \mathrm{tr}\ a\rangle_j \langle^j a - \mathrm{tr}\ FW \tag{45}$$

The gradient of the scalar differences of equation (45) is

$$\frac{\partial(\mathrm{tr}\ \Psi_j)}{\partial W_{pxk}} = \frac{\partial \langle^j a\ a\rangle_j}{\partial W} - \frac{\partial \mathrm{tr}}{\partial W}(FW) \tag{46}$$

and by equation (40) and equation (B-93)

$$\frac{\partial}{\partial W}(\mathrm{tr}\ \Psi_j) = W^T v\rangle_j \langle^j v\ 2 - F \tag{47}$$

The "expected value" over all j is

$$E\left\{\frac{\partial}{\partial W}(\mathrm{tr}\ \Psi_j)\right\} = E_j\left\{W^T v\rangle_j \langle^j v\right\} 2 - F \tag{48}$$

$$= W^T Q_{vv} 2 - F$$

Minimizing the scalar difference expression of equation (48) requires the gradient term of equation (48) to be equated to the [0] matrix.

$$W^T Q_{vv} 2 - F = [0]_{pxp} \tag{49}$$

or

$$W^T 2 = F\ Q_{vv}^{-1}\ . \tag{50}$$

The constraint of equation (29) is

$$FW = I_{pxp} \tag{51}$$

and transposing

$$W^T F^T = I \tag{52}$$

hence multiplying equation (50) by F^T

$$W^T F^T 2 = \underset{pxp}{I2} = \underset{pxp}{F\, Q_{vv}^{-1}\, F^T} \tag{53}$$

Transposing equation (50

$$\underset{kxp}{WI2} = Q_{vv}^{-1}\, F^T \tag{54}$$

and using (53)

$$\underset{(kxp)}{W}\underset{(pxp)}{(F\, Q_{vv}^{-1}\, F^T)} = \underset{kxk}{Q_{vv}^{-1}}\, \underset{kxp}{F^T} \tag{55}$$

and solving for W

$$\underset{kxp}{W} = Q_{vv}^{-1}\, F^T\, (F\, Q_{vv}^{-1}\, F^T)^{-1} \tag{56}$$

or

$$\underset{pxk}{W^T} = (F\, Q_{vv}^{-1}\, F^T)^{-1}\, F\, Q_{vv}^{-1} \;. \tag{57}$$

Utilizing the weight-matrix (56) in equation (33)

$$\langle \hat{a}_w|_j = \langle zW|_j = \langle z|_j\, Q_{vv}^{-1}\, F^T (F\, Q_{vv}^{-1}\, F^T)^{-1} \tag{58}$$

which is the weighted estimate of the parameters $\langle \alpha$.

The error in the estimate by equation (58) is

$$E_j \left\{ |\hat{a}_w\rangle_j \langle \hat{a}_w|^j \right\} = \underset{pxp}{\Sigma_{\hat{a}\hat{a}_w}} = W^T\, Q_{vv}\, W \tag{59}$$

Using (56) and (57) in equation (59)

$$\underset{p\times p}{\Sigma}_{\tilde{a}\tilde{a}_w} = (F\, Q_{vv}^{-1}\, F^T)^{-1} F\, Q_{vv}^{-1}\, Q_{vv}\, Q_{vv}^{-1}\, F^T (F\, Q_{vv}^{-1}\, F^T)^{-1} \tag{60}$$

$$\Sigma_{\tilde{a}\tilde{a}_w} = (F\, Q_{vv}^{-1}\, F^T)^{-1} \tag{61}$$

which geometrically represents the ellipsoid of uncertainty in the estimate of the p parameters.

Observe that if the noise matrix is a scalar matrix

$$E\left[\rangle_j \, {}^j\langle \right] = \underset{k\times k}{Q_{vv}} = \sigma_{vv} \underset{k\times k}{I} \tag{62}$$

where σ_{vv} is a real variable (a scalar), then the unweighted variance of the estimate of the parameters of equation (26) becomes

$$\Sigma_{\tilde{a}\tilde{a}} = (FF^T)^{-1}\, FF^T (FF^T)^{-1}\, \sigma_{vv} \tag{63}$$

$$\Sigma_{\tilde{a}\tilde{a}} = (FF^T)^{-1}\, \sigma_{vv}. \tag{64}$$

Using the "spherical" noise matrix of (64) in the weighted variance of the estimate matrix of equation ( ) yields

$$\Sigma_{\tilde{a}\tilde{a}_w} = (FF^T\, \sigma_{vv}^{-1})^{-1} = (FF^T)^{-1}\, \sigma_{vv} \tag{65}$$

Thus one does not gain anything by weighting the data when the noise is as shown in equation (62).

SECTION V.

MULTI-VARIABLE POLYNOMIAL

Many missile range data processing tasks pose the problem of simultaneously fitting time polynomials to a number of variables. For example, a three dimensional trajectory with three coordinates of position $x(t)$, $y(t)$, $z(t)$ and three coordinates of velocity $\dot{x}(t)$, $\dot{y}(t)$ and $\dot{z}(t)$ for which we wish to approximate can be expressed as

$$\begin{pmatrix} z^1(t) \\ z^2(t) \\ z^3(t) \\ z^4(t) \\ z^5(t) \\ z^6(t) \end{pmatrix} = \begin{pmatrix} x(t) \\ y(t) \\ z(t) \\ \dot{x}(t) \\ \dot{y}(t) \\ \dot{z}(t) \end{pmatrix} \qquad (1)$$

The following derivations assume q coordinates instead of six. Consider the approximating parameters for each coordinate as given by equation(IV-5) for a single variable except now the superscript 1 to q designates the coordinate, that is

$$\begin{aligned} z^1_{\cdot k} &= \langle \overset{1}{\alpha} f(p_k) \rangle + v^1_{\cdot k} = \langle \overset{1}{a} f(p_k) \rangle + e^1_{\cdot k} \\ z^2_{\cdot k} &= \langle \overset{2}{\alpha} f(p_k) \rangle + v^2_{\cdot k} = \langle \overset{2}{a} f(p_k) \rangle + e^2_{\cdot k} \qquad (2) \\ &\vdots \\ z^q_{\cdot k} &= \langle \overset{k}{\alpha} f(p_k) \rangle + v^q_{\cdot k} = \langle \overset{k}{a} f(p_k) \rangle + e^q_{\cdot k} \end{aligned}$$

Packaging the above q coordinate into a q dimensional column vector

$$\begin{pmatrix} z^1 \\ z^2 \\ \cdot \\ \cdot \\ \cdot \\ z^q \end{pmatrix}_k = \begin{pmatrix} \langle {}^{1}a \; f\rangle_k \\ \langle {}^{2}a \; f\rangle_k \\ \vdots \\ \langle {}^{q}a \; f\rangle_k \end{pmatrix} + \begin{pmatrix} v^1 \\ v^2 \\ \cdot \\ \cdot \\ \cdot \\ v^q \end{pmatrix}_k = \begin{pmatrix} \langle {}^{1}a \; f\rangle_k \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \langle {}^{q}a \; f\rangle_k \end{pmatrix} + \begin{pmatrix} e^1 \\ \cdot \\ \cdot \\ \cdot \\ e^q \end{pmatrix}_k \tag{3}$$

Factoring out the vector $f\rangle_k$

$$z(q\rangle_k = \begin{pmatrix} \langle {}^{1}a \\ \langle {}^{2}a \\ \cdot \\ \cdot \\ \cdot \\ \langle {}^{q}a \end{pmatrix} f\rangle_k + v(q\rangle_k = \begin{pmatrix} \langle {}^{1}a \\ \\ \\ \\ \langle {}^{q}a \end{pmatrix} f(q\rangle_k + (q\rangle_k \tag{4}$$

and defining the qxp matrices of parameters as

$$\underset{q\times p}{\Lambda} = \begin{pmatrix} \langle {}^{1}p)\alpha \\ \cdot \\ \cdot \\ \cdot \\ \langle {}^{q}p)\alpha \end{pmatrix} \tag{5}$$

and

$$A = \begin{pmatrix} \langle {}^{1}p)a \\ \cdot \\ \cdot \\ \cdot \\ \langle {}^{q}p)a \end{pmatrix} \tag{6}$$

hence

$$z(q\rangle_k = \underset{q\times p}{A} f(p\rangle_k + v(q\rangle_k = \underset{q\times p}{A}\, f(p\rangle_k + e(q\rangle_k \tag{7}$$

Equation (7) is the kth observation of all q variables.

If we form the data matrix for k observations

$$\left[z(q\rangle_1 ,\ z(q\rangle_2 ,\ .\ .\ .\ z(q\rangle_k \right] = \underset{q\times k}{Z} \tag{8}$$

we have the qxk matrix which equals

$$\underset{q\times k}{Z} = [A\ f(p\rangle_1 ,\ Af(p\rangle_2 ,\ .\ .\ .\ Af(p\rangle_k] + \underset{q\times k}{V} \tag{9}$$

$$= [Af\rangle_1 ,\ Af\rangle_2 ,\ .\ .\ .\ Af]_k + \underset{q\times k}{E} \tag{10}$$

Factoring out the parameter matrices

$$Z = \underset{q\times p}{A}\ \underset{p\times k}{F} + \underset{q\times k}{V} = \underset{(q\times p)}{A}\ \underset{(p\times k)}{F} + \underset{q\times k}{E} \tag{11}$$

where the pxk matrix F is

$$\underset{p\times k}{F} = [f(p\rangle_1 ,\ f(p\rangle_2 ,\ .\ .\ .\ f(p\rangle_k] \tag{12}$$

Unweighted Least Squares Estimate of the Parameter Matrix

The unweighted estimate does not require any characteristics of the noise $\underset{q\times k}{V_j}$ where we assume that there are j different sequences of noise matrices.

If equation (12) is post-multiplied by the transpose of F

$$\underset{qxk}{Z_j F^T} = \Lambda F F^T + V_j F^T \tag{13}$$

$$= A_j F F^T + E_j F^T \tag{14}$$

and the pxp matrix FF^T is full rank, then multipling by $(FF^T)^{-1}$ yields

$$\underset{qxk}{Z_j F^T} (FF^T)^{-1} = \Lambda + V_j F^T (FF^T)^{-1} \tag{15}$$

$$= A_j + E_j F^T (FF^T)^{-1} \tag{16}$$

The unweighted least squares condition is

$$\underset{(qxk)}{E_j} \underset{(kxp)}{F^T} = \underset{qxp}{[0]} \tag{17}$$

which is shown in reference (4) using partial derivatives and also shown algebraically via orthogonal projections.

Using (17) in (16)

$$\underset{qxp}{\hat{A}_j} = \underset{qxk}{Z_j} \underset{kxp}{F^T} \underset{pxp}{(FF^T)^{-1}} \tag{18}$$

The error in the estimate by equation (15) and (16) is

$$\Lambda - \hat{A}_j = \underset{qxp}{\tilde{A}_j} = -V_j F^T (FF^T)^{-1} \tag{19}$$

The transpose of (19) is

$$\underset{qxp}{\tilde{A}_j^T} = -(FF^T)^{-1} F V_j^T \tag{20}$$

The two matrix products (major and minor), (larger and smaller), (outer and inner) available are

$$\underset{(qxp)}{\tilde{A}_j} \underset{(pxq)}{\tilde{A}_j^T} = V_j F^T (FF^T)^{-2} F V_j^T \tag{21}$$

and

$$\underset{(p\times q)q\times p}{\tilde{A}_j^T \tilde{A}_j} = (FF^T)^{-1} F V_j^T V_j F^T (FF^T)^{-1} \tag{22}$$

The traces of the two are the same, that is

$$\mathrm{tr}(\tilde{A}_j \tilde{A}_j^T) = \mathrm{tr}(\tilde{A}_j^T \tilde{A}_j). \tag{23}$$

If we partition $\tilde{A}$ into p dimensional row vectors

$$\tilde{A}_j = \begin{bmatrix} {}^{1}\langle p)\tilde{a} \\ \cdot \\ \cdot \\ \cdot \\ {}^{q}\langle p)\tilde{a} \end{bmatrix}_j \tag{24}$$

and transposing

$$\underset{p\times q}{\tilde{A}_j^T} = [\tilde{a}(p\rangle_1, \ . \ . \ . \ \tilde{a}(p\rangle_q]_j \tag{25}$$

The two matrix products of equation (21) and (22) using (24) and (25) are

$$\underset{q\times q}{\tilde{A}_j \tilde{A}_j^T} = \begin{bmatrix} {}^{1}\langle p)\tilde{a} \\ \cdot \\ \cdot \\ \cdot \\ {}^{q}\langle p)\tilde{a} \end{bmatrix}_j [\tilde{a}(p\rangle_1, \ . \ . \ . \ \tilde{a}(p\rangle_q]_j \tag{26}$$

$$= \begin{bmatrix} {}^{1}\langle p)\tilde{a} \ \tilde{a}(p\rangle_1, & . \ . \ . & {}^{1}\langle p)\tilde{a} \ \tilde{a}(p\rangle_q \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ {}^{q}\langle p)\tilde{a} \ \tilde{a}(p\rangle_1, & . \ . \ . & {}^{q}\langle p)\tilde{a} \ \tilde{a}(p\rangle_q \end{bmatrix} \tag{27}$$

which is an "outer-product" of "inner-products".

The product of equation (22) is

$$\tilde{\hat{A}}_j^T \hat{A}_j = [a(p\rangle_1 , \; . \; . \; . \; a(p\rangle_{q_j}] \begin{bmatrix} {}^1\langle p)a \\ \cdot \\ \cdot \\ \cdot \\ {}^q\langle p)a \end{bmatrix} \tag{28}$$

or an "inner product" of "outer products"

$$\underset{p x p}{\tilde{A}_j^T \tilde{A}_j} = [\tilde{a}(p\rangle_1 \, {}^1\langle p)\tilde{a} + \; . \; . \; . \; \tilde{a}(p\rangle_q \, {}^q\langle p)\tilde{a}]_j \tag{29}$$

The geometrical significance of the many previous forms is obtained from the representation of the q parameter error vectors each of dimension p as a column of column vectors

$$\begin{pmatrix} \tilde{a}(p\rangle_1 \\ \cdot \\ \cdot \\ \cdot \\ \tilde{a}(p\rangle_q \end{pmatrix} = \begin{pmatrix} q\rangle_1 \\ q\rangle_2 \\ \vdots \\ q\rangle_q \end{pmatrix} - \begin{pmatrix} \hat{a}(p\rangle_1 \\ \cdot \\ \cdot \\ \cdot \\ \hat{a}(p\rangle_q \end{pmatrix} = \beta(pq\rangle \tag{30}$$

and the transpose

$$\langle qp)\beta = [{}^1\langle p)\tilde{a}, \; {}^2\langle p)\tilde{a} \; , \; . \; . \; . \; {}^q\langle p)\tilde{a}] \tag{31}$$

The dyadic product yields

$$\beta\rangle_j \, {}^j\langle\beta = \begin{bmatrix} \tilde{a}(p\rangle_1 \, {}^1\langle p)\tilde{a} & . \; . \; . & \tilde{a}(p\rangle_1 \, {}^q\langle p)\tilde{a} \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ \cdot & & \tilde{a}(p\rangle_q \, {}^q\langle p)\tilde{a} \end{bmatrix}_j \tag{32}$$

and over all j

$$E[\underset{j}{\beta}\,\beta^{1}] = \begin{bmatrix} \underset{pxp}{\Sigma}\tilde{a}\tilde{a}_{11} & \cdots & \underset{pxp}{\Sigma}\tilde{a}\tilde{a}_{1,q} \\ \vdots & & \\ & & \underset{pxp}{\Sigma}\tilde{a}\tilde{a}_{qq} \end{bmatrix} \tag{33}$$

we obtain a matrix of variance matrices.

The sums of the main diagonal matrices of equation (33) is the expected value of equation (29)

$$E_{j}\left[\underset{pxp}{A_j^{T} A_j}\right] = \underset{pxp}{\Sigma}\tilde{a}\tilde{a}_{11} + \cdots + \underset{pxp}{\Sigma}\tilde{a}\tilde{a}_{qq} \tag{34}$$

Weighted Least Squares

This section derives the sequence of weights. Consider equation (7)

$$\underset{qxp}{Z} = \underset{(qxp)}{\Lambda}\,\underset{(pxk)}{F} + \underset{qxk}{V} = AF + \underset{qxk}{E} \tag{35}$$

We seek a kxp matrix W such that post-multiplying (35)

$$\underset{pxk}{Z_j}\,\underset{kxp}{W} = \Lambda + V_j W = \hat{A}_j \tag{36}$$

where

$$\underset{(pxk)}{F}\,\underset{(kxp)}{W} = \underset{pxp}{I} \tag{37}$$

and

$$\underset{(qxk)}{E}\,\underset{(kxp)}{W} = \underset{qxp}{[0]} \tag{38}$$

then

$$\underset{q\times p}{\hat{A}_{jw}} = \underset{q\times k}{Z_j} W = A + V_j W \tag{39}$$

Factoring Z into its column space and W into its row space

$$\underset{q\times p}{\hat{A}_{jw}} = [z(q\rangle_1, z(q\rangle_2, \ldots z(q\rangle_k] \begin{bmatrix} {}^1\langle p)w \\ \cdot \\ \cdot \\ \cdot \\ {}^k\langle p)w \end{bmatrix} \tag{40}$$

or

$$\underset{q\times p}{\hat{A}_{jw}} = z(q\rangle_1 \, {}^1\langle p)w + \ldots + z(q\rangle_k \, {}^k\langle p)w \tag{41}$$

Equation (41) states that we need a sequence of p-dimension weighting row vectors so that the weight estimate of the qxp matrix of parameters A_{jw} is a linear-dyadic combination of the data vectors $z(q\rangle_k$.

When q is equal to one we see that equation (4) becomes equation (IV-34).

The error in the weighted estimate of the parameters by equation (39) is

$$A - \hat{A}_{jw} = \underset{q\times p}{\tilde{A}_{jw}} = - V_j W \tag{42}$$

The transpose of equation (42) is

$$\underset{p\times q}{\tilde{A}^T_{jw}} = -W^T V^T_j \tag{43}$$

The two matrix products are

$$\underset{(q\times p)}{\tilde{A}_{jw}} \underset{(p\times q)}{\tilde{A}^T_{jw}} = V_j W W^T V_j \tag{44}$$

and

$$\underset{(pxp)}{\tilde{A}^T_{jw}\tilde{A}_{jw}} = \underset{pxp}{W^T V^T_j V_j W} \tag{45}$$

The expected value over all j is

$$E\left\{\tilde{A}^T_{jw}\tilde{A}_{jw}\right\} = W^T E\left\{V^T_j V_j\right\} W \tag{46}$$

$$= \underset{pxk}{W^T} \underset{kxk}{Q_{VV}} \underset{kxp}{W} \tag{47}$$

As before, form a difference matrix Ψ_j between the non-linear equation (45) and the linear relation of equation (37)

$$\Psi_j = \underset{pxp}{\tilde{A}^T_j\tilde{A}_j} - \underset{pxp}{FW} \tag{48}$$

the trace of Ψ_j is

$$\mathrm{tr}\ \Psi_j = \mathrm{tr}\ \underset{pxp}{\tilde{A}^T\tilde{A}_j} - \mathrm{tr}\ FW \tag{49}$$

The gradient of the scalar of equation (49) with respect to the matrix W is by equation (B-80) and (B-93)

$$\underset{pxk}{\frac{\partial}{\partial W}(\mathrm{tr}\ \Psi_j)} = \underset{pxk}{W^T} \underset{kxk}{V^T_j V_j} 2 - \underset{pxk}{F} \tag{50}$$

The expected value over all j of equation (50) equated to the zero matrix is

$$W^T Q_{VV} 2 - \underset{pxk}{F} = \underset{pxk}{[0]} \tag{51}$$

Clearly equation (51) is the same as equation (IV-49) and the arguments of that section hold, hence

$$W = Q_{vv}^{-1} F^T (F \, Q_{vv}^{-1} F^T)^{-1} \tag{52}$$

The primary difference in the two cases is in the computation and interpretation of the Q_{vv} matrix.

Observe that V_j is a qxk matrix

$$\underset{qxk}{V_j} = [v(q_1), \; . \; . \; . \; v(q_k)] \tag{53}$$

and

$$\underset{kxq}{V_j^T} = \begin{bmatrix} {}^{1}(q)v \\ \cdot \\ \cdot \\ \cdot \\ {}^{k}(q)v \end{bmatrix}_j \tag{54}$$

and the product is

$$\underset{(kxq)(qxk)}{V^T \; V} = \begin{bmatrix} {}^{1}(q)v \\ {}^{2}(q)v \\ \cdot \\ \cdot \\ \cdot \\ {}^{k}(q)v \end{bmatrix}_j [v(q_1), \; . \; . \; . \; v(q_k)]_j \tag{55}$$

$$= \begin{bmatrix} {}^{1}(q)v \; v(q_1), & . \; . \; . & {}^{1}(q)v \; v(q_k) \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ {}^{k}(q)v \; v(q_1) & & {}^{k}(q)v \; v(q_k) \end{bmatrix} \tag{56}$$

The expected value over all j is

$$E_j \left\{ \underset{(kxq)}{V_2^T} \; \underset{(qxk)}{V_j} \right\}$$

$$= \frac{1}{j_{max}} \sum_{j=1}^{j_{max}} V_j^T \, V_j \qquad (57)$$

where j_{max} is some economical large value.

Using (57) in (39)

$$\underset{qxp}{\hat{\Lambda}_{jw}} = \underset{qxk}{Z_j} \, Q_{vv}^{-1} \, F^T (F Q_{vv}^{-1} F^T)^{-1} \qquad (58)$$

APPENDIX A - MATRIX TRACE PROPERTIES

The trace of a matrix, the trace of the product of two matrices, and the trace of a matrix-sum are useful notions to aid the development of the topics of Appendix B.

Consider a matrix $\underset{p\times m}{A}$ of p rows and m columns where $m < p$ and a matrix $\underset{m\times p}{B}$, then the product

$$\underset{p\times p}{Q_1} = \underset{p\times m}{A}\ \underset{m\times p}{B} \tag{1}$$

is a pxp matrix.

The matrices A and B can be partitioned into their row and column spaces as shown

$$A = \begin{bmatrix} a(p\rangle_1, & \ldots & a(p\rangle_m \end{bmatrix} = \begin{bmatrix} {}^{1}\langle m)a \\ \vdots \\ {}^{p}\langle m)a \end{bmatrix} \tag{2}$$

$$B = \begin{bmatrix} {}^{1}\langle p)b \\ \vdots \\ {}^{m}\langle p)b \end{bmatrix} = \begin{bmatrix} b(m\rangle_1, & \ldots & b(m\rangle_p \end{bmatrix} \tag{3}$$

The product Q can be written as a matrix of inner-products

$$Q_1 = AB = \begin{bmatrix} {}^{1}\langle m)a \\ \vdots \\ {}^{p}\langle m)a \end{bmatrix} \begin{bmatrix} b(m\rangle_1, & \ldots & b(m\rangle_p \end{bmatrix} \tag{4}$$

$$= \begin{bmatrix} {}^{1}\langle m)a\ b(m\rangle_1 & , \ldots & {}^{1}\langle m)a\ b(m\rangle_p \\ \vdots & & \\ {}^{p}\langle m)a\ b(m\rangle_1 & , \ldots & {}^{p}\langle m)a\ b(m\rangle_p \end{bmatrix} \tag{5}$$

or as a sum of dyads (outer products)

$$Q_1 = AB = \begin{bmatrix} a(p\rangle_1 \; , \; . \; . \; . \; a(p\rangle_m \end{bmatrix} \begin{bmatrix} {}^{1}\langle p)b \\ \cdot \\ \cdot \\ \cdot \\ {}^{m}\langle p)b \end{bmatrix} \tag{6}$$

$$Q_1 = a(p\rangle_1 \, {}^{1}\langle p)b + \ldots + a(p\rangle_m \, {}^{m}\langle p)b \tag{7}$$

Equation (7) expresses Q_1 as a sum of m rank-one matrices.

If we commute the product we obtain a square mxm matrix

$$\underset{m\times m}{Q_2} = \underset{m\times p}{B} \; \underset{p\times m}{A}$$

and as before Q_2 can be written as a matrix of inner-products

$$\underset{m\times m}{Q_2} = \begin{bmatrix} {}^{1}\langle p)b \\ \cdot \\ \cdot \\ \cdot \\ {}^{m}\langle p)b \end{bmatrix} \begin{bmatrix} a(p\rangle_1 \; , \; . \; . \; . \; a(p\rangle_m \end{bmatrix} \tag{8}$$

$$= \begin{bmatrix} {}^{1}\langle p)b \; a(p\rangle_1 & , \; . \; . \; . & {}^{1}\langle p)b \; a(p\rangle_m \\ {}^{m}\langle p)b \; a(p\rangle_1 & , \; . \; . \; . & {}^{m}\langle p)b \; a(p\rangle_m \end{bmatrix} \tag{9}$$

or as a sum of dyadic products

$$Q_2 = \begin{bmatrix} b(m\rangle_1 \; , \; . \; . \; . \; b(m\rangle_p \end{bmatrix} \begin{bmatrix} {}^{1}\langle m)a \\ \cdot \\ \cdot \\ \cdot \\ {}^{p}\langle m)a \end{bmatrix} \tag{10}$$

$$= b(m\rangle_1 \, {}^{1}\langle m)a + \; . \; . \; . \; + b(m\rangle_p \, {}^{p}\langle m)a \tag{11}$$

Clearly matrix multiplication is not commutative, that is

$$\underset{p\times p}{AB} \neq \underset{m\times m}{BA} \tag{12}$$

in fact the matrices are not even of the same size.

However the trace of both products are equal, that is

$$tr(\underset{p\times p}{AB}) = tr(\underset{m\times m}{BA}) \tag{13}$$

The following will clarify the above relation.

If we have a column vector $x(p\rangle$ and a row vector $\langle p)y$ of the same dimension p then the dyadic product is the square, rank-one, matrix D of p rows and p columns

$$\underset{p\times p}{D} = x(p\rangle\langle p)y = \begin{pmatrix} x^1y_1 & \cdots & x^1y_p \\ x^py_1 & \cdots & x^py_p \end{pmatrix} \tag{14}$$

If we commute the product of Equation (14) we obtain

$$\underset{1\times 1}{d} = \langle p)\ yx\ (p\rangle = y_1x^1 + y_2x^2 + \cdots + y_px^p \tag{15}$$

a scalar.

When the elements y_i and x^i are real field elements the products commute, hence

$$y_ix^i = x^iy_i \tag{16}$$

and Equation (15) [the inner product] can be written as the sum of the main diagonal terms of $x\rangle\langle y$, which the conventional definition of the trace (tr) of a matrix, hence

$$tr\left[x\rangle\langle y\right] = \langle y\ x\rangle. \tag{17}$$

The dyadic product is not as mysterious as many novices might imagine; in fact, if we write Equation (14) as

$$\begin{aligned} \underset{p\times p}{D} &= x\rangle\langle y = x\rangle\left(y_1,\ y_2\ \cdots\ y_p\right) \\ &= \left[x\rangle\, y_1,\ x\rangle y_2\ \cdots\ x\rangle y_p\right] \end{aligned} \tag{18}$$

we see that the matrix D when partitioned into its column space is a row of p parallel column vectors - all p of the vectors lie on a line, hence $x\rangle\langle y$ is said to have rank one - that is, there is only one linearly independent vector in the row "package" of column vectors.

If we take the trace of AB by Equation (5) as the sum of diagonals we obtain

$$\mathrm{tr}(AB) = {}^{1}\langle m)a\ b(m\rangle_{1} + \cdots\cdots + {}^{p}\langle m)a\ b(m\rangle_{p} \tag{12}$$

If we take the trace of dyadic sum decomposition of AB given by Equation (7) we obtain

$$\mathrm{tr}(AB) = \mathrm{tr}\left[\ a(p\rangle_{1}\ {}^{1}\langle p)b + \cdots\cdots + a(p\rangle_{m}\,{}^{m}\langle p)b\right] \tag{13}$$

The trace of a sum of matrices is the sum of the traces, hence by Equation (17)

$$\mathrm{tr}(AB) = \mathrm{tr}\ a(p\rangle_{1}\,{}^{1}\langle p)b + \mathrm{tr}\ a(p\rangle_{2}\,{}^{2}\langle p)b + \cdots\cdots + \mathrm{tr}\ a(p\rangle_{m}\,{}^{m}\langle p)b \tag{14}$$

$$\mathrm{tr}(AB) = {}^{1}\langle p)b\ a(p\rangle_{1} + {}^{2}\langle p)b\ a(p\rangle_{2} + \cdots\cdots + {}^{m}\langle p)b\ a(p\rangle_{m} \tag{15}$$

Equation (12) is a sum of p inner-products of m-dimensional vectors and Equation (15) is a sum of m inner-products of p-dimensional vector.

The sum of the main diagonal terms of Equation (9) is

$$\mathrm{tr}(BA) = {}^{1}\langle p)b\ a(p\rangle_{1} + \cdots\cdots + {}^{m}\langle p)b\ a(p\rangle_{m} \tag{16}$$

which by Equation (15) and Equation (16)

$$\mathrm{tr}(\underset{pp}{AB}) = \mathrm{tr}(\underset{m\times m}{BA}) \tag{17}$$

APPENDIX B

GRADIENTS OF SCALARS WITH RESPECT TO MATRICES

This appendix develops the gradient of a scalar-valued function with respect to a vector variable and also with respect to a matrix variable.

Case 1. $q = \langle p) a\, x (p\rangle$. Consider the scalar q which is the inner-product

$$q = \langle p) a\, x (p\rangle \tag{b-1}$$

where $\langle a$ is a fixed p dimensional row vector and $x\rangle$ is a variable column vector, or q is said to be a scalar-valued variable which is a function of the vector variable $x\rangle$.

In equation (b-1) q may be considered to have vector factors $\langle a$ and $x\rangle$.

If we have a dyad

$$Q = x\rangle\langle a \tag{b-2}$$

then it was shown in appendix A that

$$\mathrm{tr}\ Q = q \tag{b-3}$$

or

$$\mathrm{tr}\left[x\rangle\langle a\right] = \langle a\ x\rangle = q \tag{b-4}$$

The differential of equation (b-2) is

$$d\ Q = dx\rangle\langle a \tag{b-5}$$

and the trace of (b-5) is

$$\mathrm{tr}\ dQ = \mathrm{tr}\left[dx\rangle\langle a\right] = \langle a\ dx\rangle = dq \tag{b-6}$$

We may now ask to express the differential matrix dQ in terms of vector factors $dx\rangle$ and a gradient vector, that is

$$dQ = dx\rangle\left\langle\frac{\partial q}{\partial x}\right. \tag{b-7}$$

such that the trace of equation (b-7) is

$$dq = \mathrm{tr}\ dQ = \mathrm{tr}\left[dx\rangle\left\langle\frac{\partial q}{\partial p}\right.\right] = \left\langle\frac{\partial q}{\partial x}\ dx\right\rangle \tag{b-8}$$

By equation (b-7) and (b-5) we can state

$$\left\langle \frac{\partial q}{\partial x} \right. = \langle a. \tag{b-9}$$

We arrive at the result of equation (b-9) directly from (1)

$$dq = \langle a \; dx \rangle = \left\langle \frac{\partial q}{\partial x} \; dx \right\rangle \tag{b-10}$$

hence

$$\langle a = \left\langle \frac{\partial q}{\partial x} \right. . \tag{b-11}$$

Also one can consider the gradient as an operator $\left\langle \frac{\partial}{\partial p} \right.$

$$q \left\langle \frac{\partial}{\partial q} \right. = \langle a \; x \rangle \left\langle \frac{\partial}{\partial q} \right. = \langle a \left[x \rangle \left\langle \frac{\partial}{\partial q} \right. \right] \tag{b-12}$$

The dyadic-type operator

$$x \rangle \left\langle \frac{\partial}{\partial x} \right. = \begin{pmatrix} x^1 \\ x^2 \\ \cdot \\ \cdot \\ \cdot \\ x^p \end{pmatrix} \left(\frac{\partial}{\partial x_1} , \; . \; . \; . \; \frac{\partial}{\partial x_p} \right) \tag{b-13}$$

$$= \begin{bmatrix} \frac{\partial x^1}{\partial x_1} & \frac{\partial x^1}{\partial x_2}, & . \; . \; . & \frac{\partial x^1}{\partial x_p} \\ & & & \\ \frac{\partial x^p}{\partial x_1} & & & \frac{\partial x^p}{\partial x_p} \end{bmatrix} \tag{b-14}$$

when the coordinates are independent of each other, then

$$x \rangle \left\langle \frac{\partial}{\partial x} \right. = I \; . \tag{b-15}$$

Hence

$$q\left\langle \frac{\partial}{\partial x}\right. = \left\langle \frac{\partial q}{\partial x}\right. = \langle a \qquad \text{(b-16)}$$

In conclusion:

$$\text{if } q = \langle a\ x\rangle \qquad \text{(b-17)}$$

then

$$\left\langle \frac{\partial q}{\partial x}\right. = \langle a \qquad \text{(b-18)}$$

Case 2. $q = \langle x\ x\rangle$.

When q is quadratic we can write q as the trace of the dyad

$$Q = x\rangle\langle x \qquad \text{(b-19)}$$

for

$$\text{tr } Q = \text{tr } (x\rangle\langle x) = \langle x\ x\rangle = q. \qquad \text{(b-20)}$$

The differential of the dyad

$$dQ = dx\rangle\langle x + x\rangle\langle dx \qquad \text{(b-21)}$$

$$dq = \text{tr } dQ = \langle x\ dx\rangle + \langle dx\ x\rangle = 2\langle x\ dx\rangle = \left\langle \frac{\partial q}{\partial x}\ dx\right\rangle \qquad \text{(b-22)}$$

hence

$$\left\langle \frac{\partial q}{\partial x}\right. = 2\langle x \qquad \text{(b-23)}$$

Case 3. $q = \langle x\ B\ x\rangle$ (b-24)

For this case we have two different matrices

$$Q_1 = Bx\rangle\langle x \qquad \text{(b-25)}$$

and

$$Q_2 = x\rangle\langle xB \qquad \text{(b-26)}$$

which under the trace operation map down to the same scalar

$$q = \operatorname{tr} Q_1 = \operatorname{tr} Q_2 = \langle x \; B \; x \rangle \tag{b-27}$$

The differential of $Q_2 = Q$ is

$$dQ = dx\rangle\langle xB + x\rangle\langle dxB \tag{b-28}$$

The trace of (b-28) is

$$\operatorname{tr} dQ = \langle x \; B \; dx\rangle + \langle dx \; Bx\rangle \tag{b-29}$$

The differential of (b-24) is

$$dq = \langle dx \; Bx\rangle + \langle xB \; dx\rangle = \operatorname{tr} Q \tag{b-30}$$

$$dq = \langle xB^T \; dx\rangle + \langle xB \; dx\rangle \tag{b-31}$$

$$dq = \langle x \left[B + B^T\right] dx\rangle \tag{b-32}$$

we have

$$dq = \left\langle \frac{\partial q}{\partial x} \; dx \right\rangle \tag{b-33}$$

and by (b-32) and (b-33)

$$\left\langle \frac{\partial q}{\partial x} = \langle x \left[B + B^T\right] \right. \tag{b-34}$$

and for symmetric B

$$B = B^T \tag{b-35}$$

then

$$\left\langle \frac{\partial q}{\partial x} \right. = 2 \langle xB \tag{b-36}$$

Case 4. $\quad q = \langle (p) a \underset{pxm}{X} b(m)\rangle \quad$ (b-37)

The scalar q is a function of the matrix X of p-rows and m columns.

The scalar q can be written as the trace of the matrix

$$\underset{mxm}{Q} = b(m)\rangle\langle(p) a \underset{pxm}{X} \tag{b-38}$$

The differential of Q is

$$dQ = b\rangle\langle a\; dX \tag{b-39}$$

By equation (b-37), differentiating

$$dq = \langle a\; dX\; b\rangle = \mathrm{tr}\; dQ. \tag{b-40}$$

We seek a gradient matrix $\frac{\partial q}{\partial X}$ of m rows and p columns as one of the <u>factors</u> of dQ that is

$$\underset{m\times m}{dQ} = \underset{m\times p}{\frac{\partial q}{\partial X}}\;\underset{p\times m}{dX} \tag{b-41}$$

such that

$$\mathrm{tr}dQ = dq = \langle a\; dX\; b\rangle \tag{b-42}$$

Clearly by equation (b-39) and (b-41) if

$$\underset{m\times p}{\frac{\partial q}{\partial X}} = b(m\rangle\langle p)a \tag{b-43}$$

then (b-42) is satisfied.

An alternate, more direct, approach is given below. Partition X into a row of column vectors (all "contravariant" vectors), then

$$q = \langle p)a\begin{bmatrix} x(p\rangle_{1}\;, \;\ldots\; x(p\rangle_{m}\end{bmatrix}b(m\rangle \tag{b-44}$$

$$= \begin{bmatrix}\langle a\; x\rangle_{1},\; \langle a\; x\rangle_{2}\;, \;.\;.\;.\; \langle a\; x\rangle_{m}\end{bmatrix}b(m\rangle$$

$$= \begin{bmatrix}\langle a\; x\rangle_{1},\; \langle a\; x\rangle_{2}\;, \;.\;.\;.\; \langle a\; x\rangle_{m}\end{bmatrix}\begin{bmatrix} b^{1} \\ b^{2} \\ \cdot \\ \cdot \\ \cdot \\ b^{m}\end{bmatrix}$$

$$q = \langle a \; x\rangle_{1} b^{1} + \langle a \; x\rangle_{2} b^{2} + \ldots + \langle a \; x\rangle_{m} b^{m}$$

$$= q_1(x\rangle_{1}) + \quad \ldots\ldots\ldots \quad + qm(x\rangle_{m}) \tag{b-45}$$

where each q_i is a function of a single column vector $x\rangle_{i}$.

The scalar differential of q is

$$dq = \left\langle \frac{\partial q}{\partial x} \right._{1} dx\rangle_{1} + \left\langle \frac{\partial q}{\partial x} \right._{2} dx\rangle_{2} + \ldots + \left\langle \frac{\partial q}{\partial x} \right._{m} dx\rangle_{m} \tag{b-46}$$

$$dq = \left[\left\langle \frac{\partial q}{\partial x} \right._{1}, \left\langle \frac{\partial q}{\partial x} \right._{2}, \ldots \left\langle \frac{\partial q}{\partial x} \right._{m} \right] \begin{bmatrix} dx\rangle_{1} \\ dx\rangle_{2} \\ \vdots \\ dx\rangle_{m} \end{bmatrix} \tag{b-47}$$

Equation (b-47) can be written as

$$dq = tr \left\{ \begin{bmatrix} \langle \overset{1}{p}) \frac{\partial q}{\partial x} \\ \vdots \\ \langle \overset{m}{p}) \frac{\partial q}{\partial x} \end{bmatrix} \begin{bmatrix} dx(p\rangle_{1}, & \ldots & dx(p\rangle_{m} \end{bmatrix} \right\} \tag{b-48}$$

$$= tr \left\{ \begin{pmatrix} \langle \partial q \; dx\rangle_{1}, & \ldots & \langle p \; dx\rangle_{m} \\ \vdots & & \\ \left\langle \frac{\partial q}{\partial x} \; dx\right\rangle_{1} & & \left\langle \frac{\partial q}{\partial x} \; dx\right\rangle_{m} \end{pmatrix} \right\} \tag{b-49}$$

The differential of X is a row of column vectors

$$d\underset{p\times m}{X} = \left[\, dx(p\rangle_1, \; . \; . \; . \; dx(p\rangle_m \,\right] \qquad \text{(b-50)}$$

and the <u>gradient</u> <u>matrix</u> is a column of row <u>gradient</u>-<u>vectors</u>.

$$\frac{\partial q}{\partial X} = \begin{pmatrix} \langle p)^1 \dfrac{\partial q}{\partial x} \\ \cdot \\ \cdot \\ \cdot \\ \langle p)^m \dfrac{\partial q}{\partial x} \end{pmatrix} \qquad \text{(b-51)}$$

From the foregoing we write

$$dQ = \underset{m\times p}{\frac{\partial q}{\partial X}} \; dX \qquad \text{(b-52)}$$

and

$$dq = tr\; dQ = tr \left[\frac{\partial q}{\partial X} \quad dX \right] \qquad \text{(b-52)}$$

By equation (b-45), (b-46) and (b-16)

$$\langle p)_1 \frac{\partial q}{\partial x} = \left\langle \frac{\partial q_1}{\partial x} \right._{1'} = b^1 \langle a \qquad \text{(b-53)}$$

$$\cdot$$
$$\cdot$$
$$\cdot$$

$$\langle p)_m \frac{\partial q}{x} \quad \left\langle \frac{\partial q_m}{\partial x} \right._m = b^m \langle a \, .$$

Packaging the row vector gradients of (b-53) into the column of (b-51) we obtain

$$\frac{\partial q}{\partial X} = \begin{pmatrix} b^1 \langle p) a \\ b^2 \langle p) a \\ \cdot \\ \cdot \\ \cdot \\ b^m \langle p) a \end{pmatrix} = \begin{pmatrix} b^1 \\ b^2 \\ \cdot \\ \cdot \\ \cdot \\ b^m \end{pmatrix} \langle p) a \qquad \text{(b-54)}$$

or

$$\frac{\partial q}{\partial X_{mxp}} = b(m\rangle\langle p)a \tag{b-55}$$

hence in conclusion

$$\boxed{\begin{array}{l} \text{if } q = \langle p)a\, X_{pxm}\, b(m\rangle \\[2ex] \text{then } \dfrac{\partial q}{\partial X_{mxp}} = b(m\rangle\langle p)a. \end{array}} \tag{b-56}$$

<u>Case 5.</u> $q = \langle p)a\, X_{pxm}\, B_{mxp}\, a(p\rangle$ (b-57)

For this case we set

$$B_{mxp}\, a(p\rangle = b(m\rangle \tag{b-58}$$

as in equation (b-56), then

$$q = \langle a\, X_{pxm}\, b(m\rangle \tag{b-59}$$

and we obtain the case 4, hence

$$\frac{\partial q}{\partial X} = B_{mxp}\, a(p\rangle\langle p)a \tag{b-60}$$

or

$$\boxed{\begin{array}{l} \text{if } q = \langle p)a\, X_{pxm}\, B_{mxp}\, a(p\rangle \\[2ex] \text{then } \dfrac{\partial q}{\partial X_{mxp}} = B_{mxp}\, a(p\rangle\langle p)a \end{array}} \tag{b-61}$$

Case 6. $\quad q = \langle p) a \underset{p\times m}{X} \; \underset{m\times p}{X^T} b(p\rangle .$ (b-62)

This case is the matrix analog of the quadratic vector case of equation (b-24).

We can partition X into its column space and X^T into its row space and obtain

$$q = \langle p) a \left[x(p\rangle_1 , \; . \; . \; . \; x(p\rangle_m \right] \begin{bmatrix} {}^1\langle p) x \\ \cdot \\ \cdot \\ \cdot \\ {}^m\langle p) x \end{bmatrix} b(p\rangle \tag{b-63}$$

and

$$q = \langle a \left[x(p\rangle_1 \, {}^1\langle p) x + \; . \; . \; . + x(p\rangle_m \, {}^m\langle p) x \right] b(p\rangle \tag{b-64}$$

Distributing the two end vectors over the dyadic-sum decomposition of XX^T we obtain

$$q = \langle a \; x\rangle_1 \, {}^1\langle x \; b\rangle + \langle a \; x\rangle_2 \, {}^2\langle x \; b\rangle + \ldots + \langle a \; x\rangle_m \, {}^m\langle x \; b\rangle \tag{b-65}$$

Because of inner-product commutativity

$$ {}^i\langle x \; b\rangle = \langle b \; x\rangle_i \tag{b-66}$$

hence

$$q = \langle a \; x\rangle_1 \langle b \; x\rangle_1 + \; . \; . \; . + \langle a \; x\rangle_m \langle b \; x\rangle_m \tag{b-67}$$

$$= p_1(x\rangle_1) \; q_1(x\rangle_1) + \; . \; . \; . + p_m(x\rangle_m) \; q_m(x\rangle_m)$$

hence the scalar q is a sum of products of scalars $p_i \; q_i$.

We have as before

$$dq = \left[\frac{\partial q}{\partial x} , \; \frac{\partial q}{\partial x} , \; . \; . \; . \; \frac{\partial q}{\partial x} \right] \begin{bmatrix} {}^1 dx\rangle \\ \cdot \\ \cdot \\ \cdot \\ {}^m dx\rangle \end{bmatrix} = \operatorname{tr} dQ \tag{b-68}$$

where dQ is as in equation (b-41).

$$\left\langle\frac{\partial q}{\partial x}\right. = \left\langle\frac{\partial(p_i q_i)}{\partial x}\right. = q_i\left\langle\frac{\partial p_i}{\partial x}\right. + p_i \quad \frac{\partial q_i}{\partial x} \tag{b-69}$$

and

$$p_i = \langle a\; x\rangle_i \tag{b-70}$$

$$\left\langle\frac{\partial p_i}{\partial x}\right._i = \langle a \tag{b-71}$$

$$q_i = \langle p)b\; x\rangle_i \tag{b-72}$$

$$\langle p)\frac{\partial q_i}{\partial x} = \langle p)b \tag{b-73}$$

Using (b-70), (b-71), (b-72), (b-73) in (b-69)

$$\left\langle\frac{\partial q}{\partial x}\right. = q_i \langle a + p_i \langle b \tag{b-74}$$

$$= \langle x_i\; b\rangle\langle a + \langle x_i\; a\rangle\langle b$$

$$\left\langle\frac{\partial q}{\partial x}\right. = \langle x_i\left[\, b(p\rangle\langle p)a + a(p\rangle\langle p)b \,\right] \tag{b-75}$$

Packaging (b-75) into the gradient matrix of equation (b-51)

$$\underset{m\times p}{\frac{\partial q}{\partial x}} = \begin{bmatrix} \langle \overset{1}{x}\left[\, b\rangle\langle a + a\rangle\langle b \,\right] \\ \cdot \\ \cdot \\ \cdot \\ \langle \overset{m}{x}\left[\, b\rangle\langle a + a\rangle\langle b \,\right] \end{bmatrix} \tag{b-76}$$

or

$$\frac{\partial q}{\partial X} = \begin{bmatrix} \overset{1}{\langle p)}x \\ \vdots \\ \overset{m}{\langle p)}x \end{bmatrix} \begin{bmatrix} b\rangle\langle a + a\rangle\langle b \end{bmatrix} \qquad \text{(b-77)}$$

$$\frac{\partial q}{\partial X} = X^T \begin{bmatrix} b\rangle\langle a + a\rangle\langle b \end{bmatrix} \qquad \text{(b-78)}$$

In conclusion

$$\boxed{\begin{array}{l} \text{if } \quad q = \langle p)a \underset{p\times m}{X} \underset{m\times p}{X^T} b(p\rangle \\ \\ \text{then } \quad \underset{m\times p}{\frac{\partial q}{\partial X}} = \underset{m\times p}{X^T} \begin{bmatrix} b\rangle\langle a + a\rangle\langle b \end{bmatrix} \end{array}} \qquad \text{(b-79)}$$

In a similar fashion it can be shown that

$$\boxed{\begin{array}{l} \text{if } \quad q = \langle m)c \underset{m\times p}{X^T} \underset{p\times m}{X} b(m\rangle \\ \\ \text{then } \quad \underset{m\times p}{\frac{\partial q}{\partial X}} = \begin{bmatrix} c(m\rangle\langle m)b + b(m\rangle\langle c \end{bmatrix} \underset{m\times p}{X^T} \end{array}} \qquad \text{(b-80)}$$

Consider the pxp matrix L which has factors as shown

$$\underset{pxp}{L} = \underset{pxk}{B}\ \underset{kxp}{X} \qquad \text{(b-81)}$$

where X is a variable matrix.

If we factor B into its column space and X into its row space

$$L = [b(p\rangle_1, \ . \ . \ . \ b(p\rangle_k] \begin{bmatrix} {}^{1}\langle p)x \\ \vdots \\ {}^{k}\langle p)x \end{bmatrix} \qquad \text{(b-82)}$$

$$= b(p\rangle_1\ {}^{1}\langle p)x + \ . \ . \ . \ + b(p\rangle_k\ {}^{k}\langle p)x \qquad \text{(b-83)}$$

The trace of L is

$$\mathrm{tr}\ L = \ell = {}^{1}\langle x\ b\rangle_1 + \ . \ . \ . \ + {}^{k}\langle x\ b\rangle_k \qquad \text{(b-84)}$$

The differential of ( ) is

$$dL = B\ d\ X. \qquad \text{(b-85)}$$

The factors of dL can also be expressed as

$$dL = \frac{\partial(\mathrm{trL})}{\underset{pxk}{\partial X}}\ \underset{kxp}{dX} \qquad \text{(b-86)}$$

where the pxk gradient matrix is

$$\underset{pxk}{\frac{\partial \ell}{\partial X}} = \left[\frac{\partial \ell}{\partial x}(p\rangle_1,\ \frac{\partial \ell}{\partial x}(p\rangle_2,\ . \ . \ . \ \frac{\partial \ell}{\partial x}(p\rangle_k \right] \qquad \text{(b-87)}$$

The differential of Equation (   ) is

$$d(\mathrm{tr}\, L) = d\ell = d\ell_1 + \ldots + d\ell_k \tag{b-88}$$

$$= \langle \overset{1}{dx} \frac{\partial \ell}{\partial x_1} \rangle + \ldots + \langle \overset{k}{dx} \frac{\partial \ell}{\partial x_k} \rangle \tag{b-89}$$

where

$$\frac{\partial \ell}{\partial x} \rangle = \frac{\partial \ell_1}{\partial x} \rangle = \frac{\partial}{\partial x} \rangle \langle \overset{1}{x}\, b_1 \rangle = b_1 \rangle \tag{b-90}$$

$$\vdots$$

$$\frac{\partial \ell}{\partial x_k} \rangle = \frac{\partial \ell_k}{\partial x_k} \rangle = \frac{\partial}{\partial x_k} \rangle \langle \overset{k}{x}\, b_k \rangle = b_k \rangle$$

and

$$\frac{\partial \ell}{\partial X} = [b(p)_1 \rangle, \ldots b(p)_k \rangle] = \underset{pxk}{B} \tag{b-91}$$

In summary,

If

$$\underset{pxp}{L} = \underset{(pxk)}{B}\ \underset{(kxp)}{X} \tag{b-92}$$

then

$$\frac{\partial(\mathrm{tr}L)}{\underset{pxk}{\partial X}} = \underset{pxk}{B} \tag{b-93}$$

APPENDIX C

MINIMIZATION

Consider the linear surface

$$\ell = \langle b\ x \rangle \tag{1}$$

and the quadratic surface

$$q = \langle xQx \rangle \tag{2}$$

and the difference

$$q - \ell = \phi. \tag{3}$$

If ℓ is a constant, $\ell = \ell o$, then we seek a vector x that lies on the linear surface and on the quadratic surface such that difference in the linear surface and the quadratic surface is a minimum.

Differentiating

$$d\phi = dq - d\ell \tag{4}$$

and

$$d\phi = \left\langle \frac{\partial \phi}{\partial x}\ dx \right\rangle \tag{5}$$

$$= \left\langle \frac{\partial q}{\partial x}\ dx \right\rangle - \left\langle \frac{\partial \ell}{\partial x}\ dx \right\rangle \tag{6}$$

$$= \left\langle \left[\left\langle \frac{\partial q}{\partial x} - \left\langle \frac{\partial \ell}{\partial x} \right] dx \right\rangle \right. \right. \tag{7}$$

or

$$\left\langle \frac{\partial \phi}{\partial x} = \left\langle \frac{\partial q}{\partial x} - \left\langle \frac{\partial \ell}{\partial x} \right. \right. \right. \tag{8}$$

If we equate the gradient vector to zero

$$\left\langle \frac{\partial q}{\partial x} = \left\langle \frac{\partial \ell}{\partial x} \right. \right. . \tag{9}$$

By equation ( ) and equation ( )

$$2\langle x Q = \langle b \tag{10}$$

and solving for $\langle x$

$$\langle x = \frac{\langle b Q^{-1}}{2} \; . \tag{11}$$

Multiplying equation (11) by $b\rangle$ and using equation ( )

$$\langle x \; b\rangle = \frac{\langle b \; Q^{-1} \; b\rangle}{2} = \ell o \tag{12}$$

or

$$\frac{1}{2} = \frac{\ell o}{\langle b Q^{-1} b\rangle} \tag{13}$$

Using (13) in (11)

$$\langle x = \ell o \, \frac{\langle b \; Q^{-1}}{\langle b Q^{-1} b\rangle} \tag{14}$$

If

$$\ell o = 1$$

then

$$\langle x = \frac{\langle b Q^{-1}}{\langle b Q^{-1} b\rangle} \tag{15}$$

REFERENCES

1. Dwyer, Paul S., "Some Applications of Matrix Derivatives on Multi-Variate Analysis", American Statistical Association Journal, June 1967.

2. Kalman, R. E., "New Methods in Weiner Filtering Theory", Proceedings of the First Symposium in Engineering Applications of Random Function Theory and Probability. John Wiley and Sons, 1963.

3. Lee, C. K. Robert, "Optimal Estimation, Identification and Control", The M.I.T. Press, Cambridge, Mass., 1964.

4. Pappas, James S., "Application of the Kalman Filter To Sequential Optimal Parameter Estimation Via Householder's Matrix Inversion Method", R-O-S-67-1, Special Report, USATECOM, Analysis and Computation Directorate, Deputy for NRO, White Sands Missile Range, New Mexico, June 1967.

5. Pappas, J. S., Diaz, A., "A Math Model for Computing Noise Variance Matrices for a System of Radar Trackers", R-O-S-67-2, Special Report, USATECOM, Analysis and Compuation Directorate, Deputy for NRO, White Sands Missile Range, New Mexico, July 1967.

6. Scheffe, Henry, "The Analysis of Variance", John Wiley, 1959.

7. Scott, Maceo T., Rede, Edmundo, and Wygant, Marthe, "Discrete Recursive Estimation: An Optimum Automated Data Processing Technique For Multi-Radar Tracking Systems", Technical Report Nr. 5, Analysis and Computation Directorate, White Sands Missile Range, New Mexico, April 1967.

8. Friedman, B., "Principles and Techniques of Applied Mathematics", John Wiley, 1956.

DISTRIBUTION

Special Report Nr. RO-S-68-1, "A Vector Space Derivation Using Dyads of Weighted Least Squares for Correlated Noise", UNCLASSIFIED, National Range Operations, White Sands Missile Range, New Mexico 88002, June 1968.

	Number of Copies
White Sands Missile Range New Mexico (STEWS)	
National Range Operations	
Analysis and Computation Directorate (Record Copy)	1
(Ref. Copy)	1
Weapon System Simulation Division, ATTN: Mr. Pappas	69
Technical Library	10
Defense Documentation Center Cameron Station Alexandria, Virginia 22314	20

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) ANALYSIS AND COMPUTATION DIRECTORATE DEPUTY FOR NATIONAL RANGE OPERATIONS WHITE SANDS MISSILE RANGE, NEW MEXICO 88002	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP	
3. REPORT TITLE A VECTOR SPACE DERIVATION USING DYADS-OF WEIGHTED LEAST SQUARES FOR CORRELATED NOISE		
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) SPECIAL REPORT		
5. AUTHOR(S) (Last name, first name, initial) PAPPAS, JAMES S.		
6. REPORT DATE JUNE 1968	7a. TOTAL NO. OF PAGES 76	7b. NO. OF REFS 8
8a. CONTRACT OR GRANT NO. b. PROJECT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)	
c. d.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	
10. AVAILABILITY/LIMITATION NOTICES DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED		
11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY	

13. ABSTRACT

Matrix-analysis and recursive matrix computing subroutines offer hope of relieving the current computer data deluge. Classical weighted least squares for multi-variable parameter estimation in the presence of correlated noise are developed in a geometrical vector space setting. Rank-one matrices, or dyads, are used extensively, especially in obtaining gradients of traces of variance matrices.

DD FORM 1 JAN 64 1473

UNCLASSIFIED

Security Classification

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
1. Matrices 2. Weighted least squares 3. Correlated noise						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through ________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS), (S), (C), or (U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.

UNCLASSIFIED

Security Classification